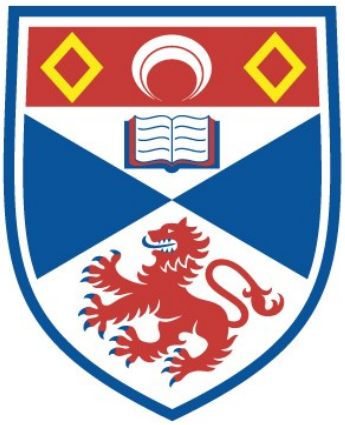
THEORETICAL INVESTIGATIONS ON SOME BIOLOGICAL SYSTEMS

Paola Scano

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



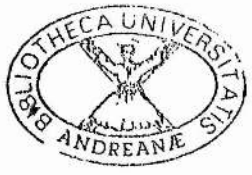
1997

Theoretical Investigations on Some Biological Systems

A thesis presented for the degree of Doctor of Philosophy in the Faculty of Science of the University of St. Andrews

By

Paola Scano



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

DECLARATION

I, Paola Scano hereby certify that this thesis has been written by me, that it is a record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Signed Date

I was admitted to the faculty of Science of the University of St. Andrews as a research student and as a candidate for the degree of Ph. D. in Theoretical Chemistry in January '88.

Signed Date

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropiate for the degree of Ph.D. in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Signed Date Sept 1st 1997

COPYRIGHT

ACKNOWLEGEMENTS

I would like to thank my supervisor Dr. Colin Thomson for giving me the opportunity to work on this project and the National Foundation for Cancer research for financial support.
I am also grateful to the University of St. Andrews and to the Universita' di Cagliari for the provision of research and library facilities.
I would like to thank Prof. Giuseppe Saba who first introduced me to the wonderful world of chemistry and computing.
A special thanks goes to Prof. Francesco P. Corongiu who first proposed that I get in touch with Dr. Colin Thomson in order to study for the Ph.D. degree in St. Andrews.
Of many people I am grateful to I would like to make a special mention to my research group and to my friends in St. Andrews that I will never forget; and last but not least to my wonderful family.

ABSTRACT

In this thesis molecular properties of different biological systems are studied by means of Quantum Mechanical methods. After a brief introduction, the basic theory of the methods here used is presented in Chapter 1.

In Chapter 2 the heats of formation and the ionization potentials of some carcinogenic compounds such as Polycyclic Aromatic Hydrocarbons are studied and compared with experimental data. The methods used are the semiempirical methods with the AM1 and PM3 Hamiltonians. The results show that the experimental heats of formation are better reproduced by PM3, while as far as the ionization potentials are concerned AM1 is slightly better than PM3.

In Chapter 3, the Lipid Peroxidation mechanism has been studied at the molecular level. The energetic details of the pathway proposed by Porter have been studied using the semiempirical AM1 and PM3 Hamiltonians. Six model compounds, which retain the same functional characteristics as the physiological fatty acid, have been used. The results are in good agreement with the experimental observations.

Characteristics and metabolism of the anticancer drug NMF (N-Methylformamide) have been studied in detail in Chapter 4. Accurate ab initio calculations have been carried out and the effect of the solvent has been taken into account. The results

represent a good example of the high quality of information that the application of ab initio methods can give when properties of molecules and energetic details of pathways are examined.

TABLE OF CONTENTS

INTRODUCTION

INTRODUCTION

The aim of this work is, to a large extent, the study of molecular properties of different biological systems by means of Quantum Mechanical methods.

The theory of Quantum Mechanics has been extensively used to deduce, calculate or extract specific molecular properties. The theoretical framework for accomplishing this goal is a solid one, and many sophisticated computational treatments have been developed from this theory. These were spurred on by dramatic improvements in computing machinery.

The study of various properties of biomolecules and the biological systems studied in this thesis are a small contribution to the bigger and ambitious project of defeating cancer.

1) CANCER RESEARCH

In the last few decades, the extensive work dedicated to the study of cancer has made it possible to shed some light on this pathological phenomenon which is today the second cause of death after cardiovascular diseases.

Cancer is a very old disease, and even the Egyptians were aware of cancer. Although Hippocrates in the 4th century B.C. described many kinds of tumours and coined the term "cancer"

(from the ancient greek word: crab), the systematic study of this disease started in the 19th century.

The origins of this disease are in part still unknown. So far the knowledge acquired has established a very close relation between life style and human behaviour and the development of cancer (see Figure 1.1). There are several types of cancer and their widely accepted classification depends on the tissue of origin.

Cancer is a general term frequently used to indicate various types of malignant neoplasms. In turn, malignant neoplasms are characterized by abnormal cell growth, in which the cells divide without normal control and can invade surrounding tissues. It is likely to recur following initial treatments, frequently metastasizes and causes the death of the patient unless adequately treated.

To treat cancer there are nowadays many different types of therapy. They range from surgery to radiotherapy, chemotherapy and even bone marrow transplantation.

For those who are not experts it is sometimes surprising to discover the amount of time and effort needed to develop a new anticancer drug. There are years and years of work under different experimental conditions, *in vitro*, *in vivo*, in mice and other animals, until the drug can be tested in patients (Clinical Trials). The last stage is the commercial development of the drug. In general, active drugs have in the past been discovered on an empirical or observational basis. Unfortunately, for most of them

we do not know exactly how they act. Some of them interact with nucleic acids and interfere with protein synthesis; their favoured targets are cells undergoing division. This explains the side effects of chemotherapy; in fact chemotherapy destroys not only cancer cells but also rapidly dividing normal body cells, such as cells lining the gastrointestinal tract, hair follicles, bone marrow cells and lymphocytes involved in the immune defense system. Chemotherapy's effectiveness depends on finding chemical agents that can interact selectively with cancer cells. Unfortunately the differences between normal cells and cancer cells are very small; the cancer cell being originally a "normal" cell of our own body, that is not anymore subject to the rules and regulations governing normal cells and the overall function of the body.

For these reasons the study at the molecular level of the anticancer drugs, of their metabolism and their interaction with cell components, and the study of the characteristics of neoplastic cells would be the logical approach to find new chemical agents effective against cancer.

2) METHODS

Since the biological systems above mentioned are quite large in terms of computational cost, we need to use approximate Quantum Mechanical methods. In the study of biological systems by these methods we move from the original system, the whole living body, to a limited number of molecules or atoms. In spite of

this "oversimplification". theoretical methods have been widely and successfully used to try to obtain a better understanding of the molecular aspects of the system.

The theoretical methods used here are all based on Quantum Chemistry. With the term Quantum Chemistry, we mean the theory of Quantum Mechanics applied to Chemistry.

Quantum Chemistry is mostly concerned with finding the best treatment to describe experimentally measured properties such as molecular geometries, conformational energies, ionization potentials, dipole moments, charge distributions etc.. Quantum Chemistry is also concerned with the study of molecular properties which are difficult to obtain experimentally. The interfacing of the results with graphic programs makes it possible to visualize the 3-D structure of molecules and molecular systems.

The two most commonly employed Quantum Chemistry methods for the study of the properties of molecules are the *ab initio* and Semiempirical Molecular Orbitals (MO) methods. These two methods are based on the solution of the time-independent Schroedinger equation by means of the Variational Principle and by the application of the Hartree-Fock method. The complexity of the equation is greatly simplified by the introduction of the Born-Oppenheimer approximation.

In *ab initio* methods, a rigorous solution of the time-independent Schroedinger equation is obtained. The main restriction of this method in accuracy is due to the truncation of the basis set. To

allow the calculations of properties of systems that are too large to be dealt with the *ab initio* techniques, approximate procedures have grown up alongside *ab initio* methodology. There are the methods which are semi-empirical in nature. In these methods, certain terms are neglected or approximated. To achieve useful accuracy, some of the terms retained are parametrized, with the parameters determined at least in part from experimental data.

3) MODELS

To study complex biological systems such as those involved in cancer, by means of theoretical studies, we need the use of well defined models.

A model according to the meaning that this word has in science, is a rather gross approximation with respect to the true system, which is in general very complex. Theoretical Chemistry is a chemical subdiscipline working, by definition, on models. The most important point to be considered is that the results of the methods above mentioned are generally valid for molecules in gas phase. Although several methods have been developed in order to simulate the effect of a solvent, we are still far from understanding the true situation of a drug or its metabolites *in vivo* , acting in the biological environment of a whole living body. Furthermore, the enzymes normally present in a biological environment can favour chemical reactions that would never

occur in gas phase. Sometimes the size of the molecules involved in a particular system to be studied are too big to be treated completely. It is then common practice to use models, for example smaller molecules which retain the same functional features of the reference molecule, but which are small enough for the relevant computations to be feasible.

The use of different models referring to the same model system is quite common; these models may select different features of the system, because there are distinct aspects of the object which deserve modelling. The choice and elaboration of a model depends on what we are seeking, always in the limits of the general criteria which assess the quality of the model.

Another point to be taken into consideration is the choice of the method to be used to study that particular system. The more sophisticated methods with no approximations and at a high level of accuracy would be the optimal choice but this is frequently prohibitive in terms of computer resources. For the major part of molecules of biological interest, *ab initio* calculations can be carried out only if powerful computers are available and the programs are able to handling the enormous number of integrals needed.

This introduction gives an overview of the aim of this thesis, a summary of the main features of the disease known as cancer, and of the quantum mechanical methods and their potentiality in the study of biological systems.

Chapter 1 deals with the basic theory of quantum mechanics relevant to this work. Chapters 2, 3 and 4 describe and report the results of the three subjects studied in this thesis; they are respectively:

1) Study of properties of some carcinogenic compounds such as Polycyclic Aromatic Hydrocarbons and similar molecules.

In the past twenty years an extensive program has been carried out to test chemical compounds for carcinogenic and other toxic properties. Several compounds have been found to cause cancer in humans; among them a very important class is that of Polycyclic Aromatic Hydrocarbons. Such compounds with a surface of more then 120Å^2, are frequently carcinogenic. Substitution of methyl groups alters the carcinogenicity. For example, Benz(a)anthracene is a weak carcinogen but 7,12-dimethylbenz(a)anthracene is potent. Linear Polycyclic Aromatic compounds are often noncarcinogenic. The active compounds interact with the nucleic acids, and form adducts with the purine or pyrimidine bases, intercalating in the DNA grooves. The major adduct from the prototype Polycyclic Aromatic Hydrocarbon, Benzo[a]pyrene, is from position 10 of the hydrocarbon to position 2 of guanine.

The properties studied here in this section (heats of formation, ionization potentials etc.) are compared with the experimental data when available. The methods used are the semiempirical AM1 and PM3 Hamiltonians. This work has been done in order to

assess the quality of these methods in predicting molecular properties for compounds of this size and having these characteristics. Unfortunately, so far, the available experimental data for molecules of this size are very limited.

2) The study at the molecular level of Lipid Peroxidation of cell membrane components.

The explicit and unequivocal relationship between Lipid Peroxidation and biochemical alterations observed in the cell leading to injury and to a variety of pathological states, including cancer or death, has been satisfactorily demonstrated. Many of the degradation by-products of Lipid Peroxidation have been shown to be extremely active, to be strong inhibitors of protein synthesis, to form DNA adducts and to be mutagenic. Establishing a certain origin for each different degradation product is definitively an important point for future research in this field.

A number of chemical and chemico-physical properties have been determined in chemical and *in vitro* peroxidizing model systems. The chemistry of Lipid Peroxidation proceeds via the reaction pathway first proposed by Porter. In this work, the energetic details of this pathway have been studied. The size of the fatty acids, intermediates and secondary products involved in the *in vivo* mechanism was prohibitive in terms of computational cost, therefore this project addressed the choice of six model compounds which retain the same functional characteristics of the physiological fatty acids involved. The calculations have been

carried out at AM1 and PM3 levels of SCF theory, along the whole pathway for all the models.

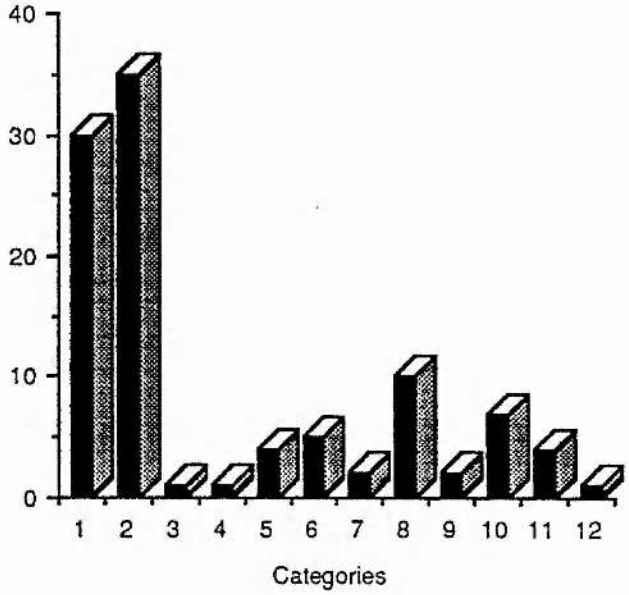
3) Study of the characteristics and metabolism of a new anticancer drug: N-Methyl Formamide (NMF).

NMF has been an agent of considerable interest to oncologists because of its broad spectrum of preclinical antitumor activity, tumor differentiating abilities and radiosensitizing and chemosensitizing properties. Analysis of the relationship between structure and inducing differentiating activity *in vitro* revealed that no specificity of structure exists to cause terminal differentiation in tumor cells. This is in contrast to the marked specificity of NMF as an antitumor agent *in vivo*. In spite of the simple structure of this molecule, the mechanism of action is still unknown. Accurate *ab initio* calculations on this molecule, were carried out with different basis sets. The potential energy surface of the proposed mechanism of action has been studied in detail to completely characterize all the stationary points: minima, intermediates and transition states. The effect of the solvent has been taken into account in order to get closer to the natural environment of the biological systems. The method used here is a continuum approach in which the solvent is treated as a polarizable dielectric, represented by a system of virtual charges on the surface of the solute cavity. The effect of these virtual charges is treated as a perturbation to the solute Hamiltonian.

Chapter 5 reports the computer facilities and the programs used in this thesis.

A brief conclusion can be found in Chapter 6.

Figure 1.1



Risk factors expressed in percentage.

Categories:

1) Tobacco; 2) Diet; 3) Additives in food; 4) Exposure to sunlight ; 5) Occupational exposure to chemicals; 6) Unknown; 7) Environmental pollution; 8) Viruses and other infections; 9) Genetic factors; 10) Sex factors; 11) Alcohol; 12) Exposure to X Ray.

(G. Bonadonna, G. Robustelli, F. Saccani; "Una sfida possibile" Rizzoli (1988))

CHAPTER 1

THEORY

1.1) THE TIME INDEPENDENT SCHROEDINGER EQUATION

The Schroedinger equation is one of the fundamental equations of Quantum Chemistry. It is usually presented as an eigenvalue problem:

$$\mathcal{H}\,\Phi = \mathcal{E}\,\Phi \tag{1.1}$$

with the operator $\mathcal{H}$ being the Hamiltonian operator and the eigenvalues $\mathcal{E}$ being the allowed energies of the system. The eigenfunctions Φ of the Hamiltonian operator are the wave functions of the system. For the non-relativistic time independent Schroedinger equation, the wave functions obtained are called stationary-state wave functions.

The wave function has a probabilistic interpretation; for a particle in a state described by $\Phi(\mathbf{r})$, where $\mathbf{r}$ are the spatial co-ordinates of the particle, $|\Phi(\mathbf{r})|^2 d\mathbf{r}$ determines the probability of finding the particle in a volume element $d\mathbf{r}$ at point $\mathbf{r}$.

For a system of N electrons and M nuclei the Hamiltonian operator, expressed in atomic units, is:

$$\mathcal{H} = -\sum_{A=1}^{M} \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 - \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{r_{ij}} + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}} \tag{1.2}$$

where:

M_A = mass of the nucleus A

∇^2 = Laplacian operator

Z_A = the atomic number of nucleus A

i,j = indices for electrons i,j=1,2...N

A,B = indices for nuclei A,B=1,2....M

R_{AB} = distance between nuclei A and B

r_{ij} = distance between electrons i and j

r_{iA} = distance between electron i and nucleus A

The first two terms represent the operators for the kinetic energy of nuclei and electrons respectively; the third term represents the coulomb attraction between electrons and nuclei. The fourth and the fifth terms are the repulsions between electrons and nuclei respectively.

The atomic unit of length is referred to as the Bohr radius, it is defined as:

$$a^0 = h^2 / (4\pi^2 m e^2) = 5.2918 \times 10^{-11} m = 0.52918 \text{ Å}$$

where h is the Planck constant, m the mass of the electron, -e the charge of the electron.

1.2) BORN-OPPENHEIMER APPROXIMATION

On account of their greater mass (the mass of a proton is about 1836 times as great as the mass of the electron) the nuclei move much more slowly than electrons. Hence it is a good approximation to neglect the first term of (1.2), considering the

electrons in a molecule moving in a potential field generated by fixed nuclei. In this approximation [1.1] the last term is a constant for a given nuclear configuration and the electronic Hamiltonian can be written as:

$$\mathcal{H}_{elect} = -\sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}} \tag{1.3}$$

The Schroedinger equation can now be written in the form:

$$\mathcal{H}_{elect}\ \Phi_R(r) = \mathcal{E}_{elect}(R)\ \Phi_R(r) \tag{1.4}$$

$\mathcal{H}_{elect}$ is now the operator for the electronic energy $\mathcal{E}_{elect}(R)$ (from now on we will drop the index **elect**). Φ depends on **r** (electron coordinates) and parametrically on **R** (nuclear coordinates), i.e. for each set of **R**'s we have a new set of Φ's.

The total energy is:

$$\mathcal{E}_{tot} = \mathcal{E}_{elect} + \sum_{A=1}^{M}\sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}} \tag{1.5}$$

The Born-Oppenheimer approximation certainly yields a more simplified form of the Hamiltonian operator but for a many-electron system is still far too complex to solve exactly.

Solutions of the electronic Schroedinger equation for many-electron systems are usually obtained by the variational method.

1.3) VARIATIONAL PRINCIPLE

If we multiply both sides of (1.1) by Φ^* and integrate over all the co-ordinates involved we obtain the energy E as the expectation value:

$$\langle E \rangle = \frac{\langle \Phi | \mathcal{H} | \Phi \rangle}{\langle \Phi | \Phi \rangle} \tag{1.6}$$

in Dirac notation (the notation for integrals used in this chapter is defined in [1.2]) .

A trial function Ψ can be used to give us an approximation to the energy:

$$\langle E(\Psi) \rangle = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} \tag{1.7}$$

where $E(\Psi)$ is the expectation value and according to the variational principle is always greater than or equal to the energy E:

$$E(\Psi) \geq E \tag{1.8}$$

The variational procedure consists in choosing a trial function of a number of variable parameters c_1 , c_2 ....., $\Psi = \Psi(c_1$, c_2 .....), and in varying these parameters until the energy reaches a minimum. Of course, the more parameters are used the more

flexible the wave function is, and the closer to E is the calculated energy.

1.4) PAULI EXCLUSION PRINCIPLE

To completely describe an electron, it is necessary to specify its spin. In this context we introduce the spin simply as being described by two spin functions α and β depending on a spin variable ω.

These two functions which correspond to spin up and spin down respectively, form a complete and orthonormal set:

$$\int \alpha^*(\omega)\alpha(\omega)d\omega = \int \beta^*(\omega)\beta(\omega)\delta\omega = 1 \qquad (1.9)$$

$$\int \alpha^*(\omega)\beta(\omega)d\omega = \int \beta^*(\omega)\alpha(\omega)\delta\omega = 0 \qquad (1.10)$$

Therefore a complete description of the electronic state is given by a function of both spatial (r) and spin coordinates (ω) :

$$\Psi(x) = \Psi(r,\omega) \qquad (1.11)$$

The Pauli Exclusion Principle [1.3], which states that the electronic wave functions must be antisymmetric under the interchange of any two electrons :

$$\Psi(x_1,x_2.....x_ix_j.....x_N) = -\Psi(x_1,x_2.....x_jx_i.....x_N) \qquad (1.12)$$

must be obeyed by valid many electron wave function.

1.5) SPIN ORBITALS AND SLATER DETERMINANTS

The wave function for one electron is called an orbital. A spatial orbital depends on the position vector $\mathbf{r}$: $\psi(\mathbf{r})$. If we introduce the spin functions, we have spin orbitals: $\chi(\mathbf{x})$:

$$\chi(\mathbf{x}) = \begin{cases} \psi(\mathbf{r})\alpha(\omega) \\ \text{or} \\ \psi(\mathbf{r})\beta(\omega) \end{cases} \qquad (1.13)$$

where $\mathbf{x}$ represents total space and spin coordinates.

The spin orbitals can be chosen to form an orthonormal set:

$$\int \chi_i(\mathbf{x})\chi_j(\mathbf{x})\, d\mathbf{x} = \delta_{ij} \qquad \delta_{ij} \begin{cases} =0 & i \neq j \\ =1 & i = j \end{cases} \qquad (1.14)$$

Since two electrons cannot occupy the same spin orbital, the wave function for N electrons is well described by the short hand notation of a determinant called a Slater determinant:

$$\Psi(\mathbf{x}_1,\mathbf{x}_2.....\mathbf{x}_i\mathbf{x}_j.....\mathbf{x}_N) = (N!)^{-1/2} \begin{vmatrix} \chi_i(\mathbf{x}_1)\ \chi_j(\mathbf{x}_1).\ .....\chi_k(\mathbf{x}_1) \\ \chi_i(\mathbf{x}_2)\ \chi_j(\mathbf{x}_2).....\chi_k(\mathbf{x}_2) \\ \\ \\ \chi_i(\mathbf{x}_N)\ \chi_j(\mathbf{x}_N).....\chi_k(\mathbf{x}_N) \end{vmatrix} \qquad (1.15)$$

$(N)^{-1/2}$ is the normalization factor.

A short hand notation for equation (1.15), which includes the normalization factor and only shows the diagonal elements of the determinant, can be written in Dirac notation as:

$$\Psi = |\chi_i(x_1)\,\chi_j(x_2).....\chi_k(x_N)\rangle \qquad (1.16)$$

One of the methods for determining the spatial part of the spin orbitals in (1.16) is the Hartree-Fock self consistent method [1.4, 1.5].

In the Restricted Hartree-Fock formalism which is used for closed shell systems, we constrain the spatial part of the spin orbitals with α and β spin to be the same $\chi_{2i-1}=\psi_i\alpha=\psi_i$, $\chi_{2i}=\psi_i\beta=\bar{\psi}_i$; then we can write the equation (1.15) as:

$$\Psi = |\psi_1,\bar{\psi}_1,\psi_2,\bar{\psi}_2.....\psi_{N/2},\bar{\psi}_{N/2}\rangle \qquad (1.17)$$

For a molecule ψ_1,ψ_2.....are called molecular orbitals.

1.6) COULOMB AND EXCHANGE OPERATORS

Using the above formalism for the expansion of Ψ, the energy for a closed shell molecule can be written as:

$$E=\langle\Psi|H|\Psi\rangle=2\sum_{i}^{N/2} h_{ii}+\sum_{i}^{N/2}\sum_{j}^{N/2}(2J_{ij}-K_{ij}) \qquad (1.18)$$

where:

$$h_{ii} = \int \psi_i^*(i)\,[-1/2\,\nabla_i^2 - \sum Z_A/r_{iA}\,]\,\psi_i(i)\,d\tau_i \qquad (1.19)$$

$$J_{ij} = \int \psi_i^*(i)\psi_j^*(j)\ |1/r_{ij}|\ \psi_i(i)\psi_j(j)\ d\tau_i d\tau_j \qquad (1.20)$$

$$K_{ij} = \int \psi_i^*(i)\psi_j^*(j)\ |1/r_{ij}|\ \psi_i(j)\psi_j(i)\ d\tau_i d\tau_j \qquad (1.21)$$

where the one-electron integral h_{ii} represents the energy of an electron in a ψ_i molecular orbital in the field of bare nuclei. h_i is the core operator describing the kinetic energy of electrons and nucleus-electron attraction terms.

The two-electron integrals J_{ij} represent the interaction between the charge distributions $|\psi_i(i)|^2$ and $|\psi_j(j)|^2$. They give the value that the total electron-electron repulsion would have if all electrons moved independently in the orbitals to which they are assigned.

K_{ij} arises from the antisymmetric form of the single determinant wave function. It has no classical interpretation. It describes the exchange repulsion between electrons with parallel spin.

The Coulomb operator $J_j(1)$ and the exchange operator $K_j(1)$ can be defined as:

$$J_j(1)\ \psi_i(1) = [\int \psi_j(2)\ |1/r_{ij}|\ \psi_j(2)\ d\tau_2\ ]\psi_i(1) \qquad (1.22)$$

$$K_j(1)\ \psi_i(1) = [\int \psi_j(2)\ |1/r_{ij}|\ \psi_i(2)\ d\tau_2\ ]\psi_j(1) \qquad (1.23)$$

1.7) HARTREE-FOCK EQUATION FOR MOLECULES

What we must do, now according to the variational method, is to minimize E with respect to the molecular orbitals.

Rewriting Eq. (1.18) as:

$$E = 2 \sum_{i=1}^{N/2} \int \psi_i^* [h + \sum_{j}^{N/2} (2J_j - K_j)] \psi_i \, dr \qquad (1.24)$$

we require E to be stationary i.e.:

$$\delta E = 0 \qquad (1.25)$$

subject to the constraint that the molecular orbitals remain orthonormal:

$$\int \psi_i^* \psi_j \, dr = \delta_{ij} \qquad (1.26)$$

where δ_{ij} is the Kronecker δ .

Within the functional variational technique, the molecular orbitals are varied by an arbitrary infinitesimal amount i.e.:

$$\psi_i = \psi_i + \delta \psi_i \qquad (1.27)$$

Consider an arbitrary variation in E caused by the infinitesimal variation $\delta \psi_i$, then from (1.24)

$$\delta E = 2 \sum_{i=1}^{N/2} \int \delta \psi_i^* [h + \sum_{j}^{N/2} (2J_j - K_j)] \psi_i \, dr$$

$$+ 2 \sum_{i=1}^{N/2} \int \delta \psi_i [h^* + \sum_{j}^{N/2} (2J_j^* - K_j^*)] \psi_i^* \, dr \qquad (1.28)$$

defining the Fock operator ($\mathbf{F}$) as:

$$\mathbf{F} = [h + \sum_{j}^{N/2} (2J_j - K_j)] \qquad (1.29)$$

then:

$$\delta E = 2 \sum_{i=1}^{N/2} \int \delta \psi_i^* F \psi_i \, dr + 2 \sum_{i=1}^{N/2} \int \delta \psi_i F^* \psi_i^* \, dr \qquad (1.30)$$

A standard mathematical technique for the solution of Eq. (1.30) is the use of undetermined Lagrangian multipiers. The introduction of the Lagrangian multipliers $-2\varepsilon_{ji}$ into (1.30) leads to:

$$\mathbf{F} \psi_i = \sum_{j}^{N/2} \psi_j \varepsilon_{ji} \qquad (1.31)$$

Since the ε matrix is Hermitian matrix Eq. (1.31) may be diagonalized leading to N/2 equation of the form:

$$\mathbf{F} \psi_i = \psi_i \varepsilon_i \qquad (1.32)$$

where ε_i are the orbital energies of the molecular orbitals ψ_i and eigenvalues of the operator $\mathbf{F}$. $\mathbf{F}$ is one-electron Hamiltonian and has been defined in (1.29).

Equations (1.32) are the Hartree-Fock equations for a closed shell. They are pseudo-eigenvalue equations since F depends on ψ_i, then they need to be solved in an iterative manner.

1.8) THE LCAO METHOD AND ROOTHAAN-HALL EQUATIONS [1.4, 1.5]

To solve the Hartree-Fock equations (1.32)' for molecules we expand the molecular orbitals in terms of a normalized but non orthogonal set of known spatial atomic orbitals { $\phi\mu(\mathbf{r})$, μ=1,2.....k }

$$\psi_i = \sum c_{\mu i}\, \phi\mu \qquad i=1,2.....k \qquad (1.33)$$

Introducing the linear expansion (1.33), the integro-differential equations (1.32) can be converted to a set of algebraic equations and we can solve the Hartree-Fock equations by standard matrix techniques. This procedure is called : Linear Combination of Atomic Orbitals (LCAO).

Substituting the linear expansion (1.33) in (1.32), multiplying both sides by $\phi\mu^*$ and integrating leads to :

$$\sum_{\nu} c_{\nu i} \int dr\ \phi\mu^*(1)\ F(1)\ \phi\nu(1) = \varepsilon_i \sum_{\nu} c_{\nu i} \int dr\ \phi\mu^*(1)\ \phi\nu(1) =$$

$$= \sum_{\nu} c_{\nu i}\, F_{\mu\nu} = \varepsilon_i \sum_{\nu} c_{\nu i}\, S_{\mu\nu} \qquad (1.34)$$

where:

$$S_{\mu\nu} = \int dr \phi\mu^*(1)\ \phi\nu(1) \qquad (1.35)$$

are the elements of the overlap matrix **S**;

and:

$$F_{\mu\nu} = \int dr \phi\mu^*(1)\ F(1)\ \phi\nu(1) \qquad (1.36)$$

are the elements of the Fock matrix **F**.

The Hartree-Fock equations are now:

$$\sum_{\nu} F_{\mu\nu}c_{\nu i} = \varepsilon_i \sum_{\nu} S_{\mu\nu}c_{\nu i} \qquad i=1,2.....k \qquad (1.37)$$

or in matrix form:

$$\mathbf{F\,c = \varepsilon\,S\,c} \qquad (1.38)$$

The equations (1.38) are called Roothaan-Hall [1.4, 1.5] equations.

1.9) THE DENSITY MATRIX

If we have a closed-shell molecule described by a single determinant wave function with each occupied orbital ψ_i containing two electrons, then the total charge density is:

$$\rho(r) = 2\sum_{i}^{N/2} \psi_i^*(r)\psi_i(r) = 2\sum_{i}^{N/2} |\psi_i(r)|^2 \qquad (1.39)$$

Or expanding:

$$\rho(r) = \sum_{i}^{N/2} \sum_{\nu} c_{\nu i}^*\, \phi\nu^*(r) \sum_{\mu} c_{\mu i}\, \phi\mu(r) =$$

$$= \sum_{\mu\nu} [\, 2\sum_{i}^{N/2} c_{\mu i}\, c_{\nu i}^*]\, \phi\mu(r)\, \phi\nu^*(r) =$$

$$= \sum_{\mu\nu} P_{\mu\nu}\, \phi\mu(r)\, \phi\nu^*(r) \qquad (1.40)$$

where:

$$P_{\mu\nu} = 2 \sum_{i}^{N/2} c_{\mu i} c_{\nu i}^{*} \qquad (1.41)$$

are the elements of the density matrix $\mathbf{P}$.

The elements of the Fock matrix (1.36) can be written as:

$$F_{\mu\nu} = dr_1 \int \phi\mu^{*}(1)\, h(1)\, \phi\nu(1) + \sum_{i}^{N/2} \int dr_1\, \phi\mu^{*}(1)\, [\, 2J_i(1)-K_i(1)]\, \phi\nu(1) \qquad (1.42)$$

where

$$H_{\mu\nu}^{core} = \int dr_1\, \phi\mu^{*}(1)\, h(1)\, \phi\nu(1) \qquad (1.43)$$

are elements of the core hamiltonian matrix. The one electron operator $h(1)$ describes the kinetic and nuclear attraction of an electron.

Expanding the molecular orbitals the expression (1.42) can be written as:

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \sum_{i}^{N/2} \sum_{\lambda\sigma} c_{\lambda i}\, c_{\sigma i}^{*}[\, 2\,(\mu\nu|\sigma\lambda) - (\mu\lambda|\sigma\nu)] =$$

$$= H_{\mu\nu}^{core} + \sum_{\lambda\sigma} P_{\lambda\sigma}\, [\, (\mu\nu|\sigma\lambda) - 1/2\, (\mu\lambda|\sigma\nu)] = H_{\mu\nu}^{core} + G_{\mu\nu} \qquad (1.44)$$

where $G_{\mu\nu}$ is the two-electron part of the Fock matrix.

We can see that $\mathbf{F}$ is a function of the expansion coefficent matrix $\mathbf{c}$: $\mathbf{F}=\mathbf{F}(\mathbf{c})$ and the equation (1.38) must be solved in an iterative manner.

Another problem to solve is to orthogonalize the set of atomic orbitals $\{\phi\mu\}$. This is usually done by diagonalizing the overlap matrix $\mathbf{S}$ to obtain a transformation matrix $\mathbf{X}$.

1.10) SCF PROCEDURE

Since $\mathbf{F}$ depends on $\mathbf{c}$, as we have pointed out, the equation (1.38) must be solved iteratively. This is done first by making a guess at the density matrix and then calculating $\mathbf{F}$ and $\mathbf{c}$. A new density matrix $\mathbf{P}$ is built from $\mathbf{c}$ and so on until we reach self consistency i.e. the density matrix calculated from the last cycle is the same as the previous one within specified criteria.
The energy is calculated as :

$$E = 1/2 \sum_{\mu} \sum_{\nu} P_{\mu\nu} (H_{\mu\nu}^{core} + F_{\mu\nu}) \qquad (1.45)$$

1.11) UNRESTRICTED HARTREE-FOCK FORMALISM

In the Restricted Hartree-Fock formalism, the restricted Slater determinant has been described as formed from spin orbitals whose spatial part are restricted to be the same for α and β spins (i.e $\{\chi_i\}=\{\psi_i\alpha,\psi_i\beta\}$). Each molecular orbital is either empty, or doubly occupied. Open-shell systems cannot be treated directly in this way, because the orbital occupation is not restricted to zero or two electrons, but one or more orbitals can be singly occupied by an electron which has either α or β spin. If we consider now a

wave function made up of N^{α} spin-orbitals with spin α, and N^{β} spin-orbitals with spin β without the requirement that a given molecular orbital be occupied with two electrons of opposite spin, but rather that spin up electrons occupy spatial orbitals different than those occupied by by the spin down electrons, we have what is called a spin Unrestricted Hartree-Fock wavefunction [1.2,1.6].

The energy expression for such a wavefunction is:

$$E = \sum_{i}^{\alpha+\beta} h_i + \frac{1}{2}\sum_{i,j}^{\alpha+\beta} J_{i,j} - \frac{1}{2}\sum_{i,j}^{\alpha} K_{i,j} - \frac{1}{2}\sum_{i,j}^{\beta} K_{i,j} \qquad (1.46)$$

Minimizing the energy with respect to the molecular orbital coefficents $\mathbf{c}^{\alpha}$ and $\mathbf{c}^{\beta}$ we obtain the so-called unrestricted Pople-Nesbet equations [1.6] :

$$\mathbf{F}^{\alpha}\,\mathbf{c}^{\alpha} = \mathbf{e}^{\alpha}\,\mathbf{S}\,\mathbf{c}^{\alpha} \qquad (1.47)$$

$$\mathbf{F}^{\beta}\,\mathbf{c}^{\beta} = \mathbf{e}^{\beta}\,\mathbf{S}\,\mathbf{c}^{\beta} \qquad (1.48)$$

These equations can be solved in manner similar to the way equation 1.38 is solved, except that must be solved simultaneously since F^{α} and F^{β} depend on both $\mathbf{c}^{\alpha}$ and $\mathbf{c}^{\beta}$.

The elements of the Fock matrices can be written

$$F^{\alpha}_{\mu\nu} = H_{\mu\nu}^{core} + \sum_{\lambda\sigma} [(P^{\alpha}_{\lambda\sigma} + P^{\beta}_{\lambda\sigma})\,(\mu\nu|\sigma\lambda) - P^{\alpha}_{\lambda\sigma}(\mu\lambda|\sigma\nu)] \qquad (1.49)$$

$$F^{\beta}_{\mu\nu} = H_{\mu\nu}^{core} + \sum_{\lambda\sigma} [(P^{\alpha}_{\lambda\sigma} + P^{\beta}_{\lambda\sigma})\,(\mu\nu|\sigma\lambda) - P^{\beta}_{\lambda\sigma}(\mu\lambda|\sigma\nu)] \qquad (1.50)$$

in which:

$$P^{\alpha}_{\mu\nu} = \sum_{i}^{N^{\alpha}} c^{\alpha}_{\mu i} c^{\alpha}_{\nu i} \qquad (1.51)$$

$$P^{\beta}_{\mu\nu} = \sum_{i}^{N^{\beta}} c^{\beta}_{\mu i} c^{\beta}_{\nu i} \qquad (1.52)$$

are the elements of the density matrices and $P^{\alpha} + P^{\beta} = P^{tot}$

The use of different orbitals for different spins it is a more correct description of open-shell molecules and it gives always lower energy than the one obtained under the restricted determinant formalism. It should be pointed out at this point that when $N^{\alpha} = N^{\beta}$ there are two possible solutions, one in which $P^{\alpha} = P^{\beta} = 1/2 P^{tot}$ and the Pople-Nesbet equations reduce to the Roothaan equations: the second in which $P^{\alpha} \neq P^{\beta}$ yields the unrestricted solution of lower energy.

Unfortunately, unrestricted wave functions are not eigenfunctions of the spin operator S^2 and they do not correspond a pure spin state. The expectation value of S^2 is always too large because the contaminants have larger values of S.

1.12) RHF HALF-ELECTRON FORMALISM

Within a modified closed-shell method, the unpaired electron of a radical is replaced by two half-electrons of opposite

spin. This method is called RHF-HE (Restricted Hartree-Fock Half Electron) [1.7] and allows calculations on open-shell systems by single Slater determinants.

If we denote by ψ_i the doubly occupied MO's, and by ψ_o the singly occupied one the total energy is expressed as:

$$E = 2\sum_{i}^{N/2} h_i + h_0 + \sum_{i,j}^{N/2} (2J_{ij} - K_{ij}) + \sum_{i}^{N/2} (2J_{i0} - K_{i0}) - 1/4\, J_{00} \qquad (1.53)$$

Where the last term is a correction apported to the usual RHF energy.
If we have only one unpaired electron the orbital ψ_o is occupied by half an α electron with spin and half an electron with β spin.

A radical is then represented symmetrically by a single determinant where the unpaired electron is replaced by two "half electrons" of opposite spin. A half electron in this sense would be defined as an imaginary particle with half the charge and twice the mass of an electron.

1.13) BASIS SETS

The sets of atomic orbitals, $\{ \phi\mu \}$, used to determine the unknown molecular orbitals in the LCAO method are called a basis set.

The basis sets that most resemble the hydrogen atomic orbitals are those comprising Slater type functions. These differ from the hydrogen like orbitals in that they have no radial nodes. The Slater type orbitals (STO) have the form:

$$r^{n-1}\, e^{-\zeta r}\, Y_{lm}(\theta,\phi) \qquad (1.54)$$

where n ,l , m are the principal, azimuthal and magnetic quantum numbers respectively; r, θ and ϕ polar co-ordinates, ζ is the Slater exponent. From a computational point of view the integrals over these functions are quite difficult to evaluate, particularly in computing the three and four center two-electron integrals [1.8]. Hence, Slater orbitals have been generally replaced by Gaussian Type Functions (GTF) [1.9] of the form:

$$x^{l}\, y^{m}\, z^{n}\, \exp -\alpha r^2 \qquad (1.55)$$

where α is the Gaussian exponent. One can define $L=l+m+n$, and refers to L=0 functions as 's' functions, L=1 as 'p' functions, L=2 as 'd' etc.. With the Gaussian functions, the computational difficulties of the STOs have been overcome, because the product of two gaussians is also a gaussian with its center lying between the centres of the two original gaussian functions. The principal disadvantages of GTFs are their smooth behaviour (lack of cusp) at the nucleus, and their too rapid decrease at large distances. This improper behaviour at the shorter and longer distances is corrected by the use of linear combinations of GTF's, ϕ_k, known as contracted functions as follows:

$$\phi_\mu' = \sum^{N} d_k \phi_k \qquad (1.56)$$

where the coefficents, d_k, are fixed, and N´is the degree of contraction [1.10]. The ϕk are the primitive gaussians.

The basis sets used in this work are of the gaussian-type developed by Pople and co-workers [1.11, 1.12, 1.13, 1.14]. A characteristic of these basis sets is that 's' and 'p' functions with the same quantum number n share the same exponent. They range from the small STO-3G [1.11] minimal basis set up to extended basis sets. In the STO-3G basis set, each atomic orbital is represented by a single contracted function of three primitives.

The split-valence basis sets, for example 3-21G [1.12] and 6-31G [1.13], make the orbitals more flexible by splitting them in an inner and an outer part. In the 6-31G basis set, the inner electrons are described by one contracted function which is a linear combination of 6 primitives, while the valence orbitals are described by two basis functions, one of which is the contraction of 3 primitives and the other is a single primitive. Basis functions with higher L may also be added to the expansion. These functions, called "polarization" functions, are d functions for the second row atoms and p functions for hydrogen atoms [1.14].

The computer time required for ab initio calculations is so strongly dependent on the number of basis function that relatively minor changes in the basis can render a calculation intractable. At the Hartree-Fock level, the numerous electron-

electron repulsion integrals, which are necessary to solve the SCF equations, increase as the fourth power of the number of basis functions.

Replacement of linear combinations of Gaussian primitive basis functions with variational coefficents, with a smaller number of contracted basis functions with fixed coefficents drastically reduces the amount of storage required to hold the two-electron repulsion integrals as well as the time required for the SCF and electron correlation portions of the calculations.

1.14) SEMIEMPIRICAL METHODS

As stated above, the most difficult and time-consuming step in the exact solution of Roothaan equations is the evaluation and handling of a large number of electron repulsion integrals.

To allow calculations on large molecules at a reasonable cost in terms of computer resources, a variety of approximations have been made, primarily to reduce the number of integrals which have to be calculated. These approaches are called semi-empirical methods. The error introduced using semiempirical models is compensated through the use of parameters determined by comparison of calculations with experiment and, in some cases, with ab initio calculations. A wide variety of such methods are available; in this thesis AM1 (Austin Model 1) [1.15] and the more recent developed MNDO-PM3 (Modified Neglect of Diatomic

Overlap, Parametric Method 3) [1.16, 1.17] have been used. They are based on the NDDO (Neglect of Diatomic Differential Overlap) [1.18, 1.19] approximation and the parameters were optimized to reproduce mainly: 1) heats of formation, 2) dipole moments, 3) ionization potentials, 4) geometries.

Basic Approximations

Within these methods, the approximation is made that the inner electrons are so close to the nuclei, as compared to valence electrons, that they can be considered part of the unpolarizable core. The charge of the core is then set equal to that of the nucleus minus that of the inner electrons. All two-electron integrals which depend on the overlapping charge densities of basis orbitals on different atoms are neglected. Thus $(\mu\nu|\lambda\sigma)$ is zero unless μ, ν belong to the same atom A and λ, σ are centered on a common atom B (which could, of course, be the same atom as A).

The theoretical framework for AM1 and PM3 is based on the NDDO method:

With these approximations the Roothaan-Hall equations become:

$$\sum_{\nu} (F_{\mu\nu} - \varepsilon_i \delta_{\mu\nu}) c_{\nu i} = 0 \qquad (1.57)$$

where $\delta_{\mu\nu}$ is the Kronecker δ.

The elements of the Fock matrix with μ and ν on atom A and λ and σ on atom B are:

diagonal elements:

$$F_{\mu\mu} = U_{\mu\mu} + \sum_{B} V_{\mu\mu,B} + \sum_{\nu}^{A} P_{\nu\nu} [(\mu\mu|\nu\nu) - 1/2(\mu\nu|\mu\nu)] + \sum_{B} \sum_{\lambda,\sigma}^{B} P_{\lambda\sigma}(\mu\mu|\lambda\sigma) \quad (1.58)$$

off diagonal :

$$F_{\mu\nu} = \sum_{B} V_{\mu\nu,B} + 1/2\, P_{\mu\nu} [3(\mu\nu|\mu\nu) - (\mu\mu|\nu\nu)] + \sum_{B} \sum_{\lambda,\sigma}^{B} P_{\lambda\sigma}(\mu\nu|\lambda\sigma) \quad (1.59)$$

$$F_{\mu\lambda} = H_{\mu\lambda} - 1/2 \sum_{\nu}^{A} \sum_{\sigma}^{B} P_{\nu\sigma} (\mu\nu|\lambda\sigma) \quad (1.60)$$

where:

1) $U_{\mu\,\mu}$ is the one-center one-electron energy which represents the sum of the kinetic energy of an electron in ϕ_μ at atom A, and its potential energy due to the attraction by the core of atom A

2) $V_{\mu\nu,B}$ is the two-center one-electron attraction energy between an electron in the distribution $\phi_\mu \phi_\nu$ at atom A and the core of atom B.

3) $g_{\mu\nu} = (\mu\mu|\nu\nu)$ are the one-center two-electron Coulomb integrals.

4) $h_{\mu\nu} = (\mu\nu|\mu\nu)$ are the one-center two-electron exchange integrals.

5) $H_{\lambda\mu}$ are the off-diagonal core matrix elements sometimes called resonance integrals.

The total energy E_{tot} is the sum of the electronic energy E_{elect} and the repulsion energy $E_{AB}{}^{core}$ between cores of atom A and B:

$$E_{tot} = E_{elect} + \sum \sum E_{AB}^{core} \qquad (1.61)$$

The heat of formation ΔH_f^{mol} is:

$$\Delta H_f^{mol} = E_{tot} - \sum E_{el}^{A} + \sum \Delta H_f^{A} \qquad (1.62)$$

the last two terms are the electronic energy of the atoms and the experimental heats of formation of the atoms.

Within the AM1 and PM3 formalisms, not all the various terms in the Fock matrix and the repulsions E_{AB}^{core} are treated analytically. Experimental values and adjustable parameters are introduced in order to compensate for the errors due to the simplifying assumptions of the NDDO scheme.

The resonance integrals, $H_{\mu\lambda}$, are approximated using the overlap integral, $S_{\mu\lambda}$. Within AM1 and PM3, $H_{\mu\lambda}$ is approximated by:

$$H_{\mu\lambda} = S_{\mu\lambda}\ 1/2\ (\beta_\mu + \beta_\lambda) \qquad (1.63)$$

where β_μ and β_λ are parameters

In AM1 and PM3 the one-center two-electrons integrals $g_{\mu\nu}$, $h_{\mu\nu}$ are derived from experimental data on isolated atoms, mostly taken from Oleari's work [1.20]

The derivation of the repulsion integrals $g_{\mu\nu}$ and $h_{\mu\nu}$ from experimental data, in some way, includes correlation effects that are formally neglected in the MO approach.

The core-core repulsion terms E_{AB}^{core} and the core-electron attraction terms $V_{\mu\nu,B}$ were modified in the MNDO (Modified

Neglect of Diatomic Overlap) [1.21] method (on which AM1 and PM3 are based) but it gave rise to a tendency to overestimate repulsion between atoms at approximately their Van der Waal distance.

In AM1 and PM3, the MNDO modification to the core-core terms is retained with the addition of spherical gaussian functions and three adjustable parameters. PM3 has two gaussians per atom, while AM1 has between two and four.

In MNDO we have 12 variables per element, 7 of which were optimized, while the remaining 5 one-center two-electron repulsion integrals are taken from atomic spectra.

AM1 is a reparametrized version of MNDO, where the core-core repulsion is corrected by the introduction of up to 4 Gaussians for each atom each with 3 parameters, then we have a second set of 12 parameters. Each atom is described by 10 to 19 parameters. By 1985 four elements were parametrized, C, H, N, and O. Today seventeen elements including F, Cl, Br, I, Si, Al, P, S, B, Zn, Hg, Ge, and Sn, have been included in the parameter base of AM1.

PM3 differs from AM1 in that the former treats the one-center, two-electron integrals as pure parameters, as opposed to being derived from atomic spectroscopy. In PM3 all quantities that enter the Fock matrix and the total energy expression have been treated as pure parameters. To accomplish this large task of optimizing parameters a new procedure has been introduced.

Thirteen elements more than AM1 have been introduced. For each element 18 parameters are used, for H atom 12.

1.15) IONIZATION POTENTIAL

The eigenvalues of :

$$\mathbf{F}\,\chi_i = \varepsilon_i\,\chi_i \qquad (1.64)$$

are called orbital energies. The value of the orbital energy for electron i is taken as a measure of the binding energy of the ith electron to the molecule. The negative of the energy of the highest occupied molecular orbital (HOMO) is, following Koopmans' theorem, the ionization potential of the molecule [1.22]. The main approximation involved is that no allowance is made for the relaxation of the orbitals upon the loss of the electron in HOMO (i.e. there is no reorganization of the other electrons).

1.16) POTENTIAL ENERGY SURFACES

The course of chemical reactions is determined by the shape of the potential energy surface comprising reactants and products, and intermediates. This is a function of 3N-6 variables, where N is the number of the atoms involved.

Along the path of minimum potential energy we consider two different types of stationary points: the minima corresponding to the reactants and the products and the maxima corresponding to

the transition states. At these points, the gradient of the potential energy E vanishes; that is :

$$g_i = \delta E/\delta x_i = 0 \qquad i=1,2.....3N-6 \quad (1.65)$$

where x_i is some set of 3N-6 internal coordinates for a molecule containing N atoms.

The stationary points are now characterized by the Hessian force constant matrix $\mathbf{H}$:

$$\mathbf{H} = \delta^2 E/\delta \mathbf{x}^2 \qquad (1.66)$$

which has, in the case of transition states, one and only one negative eigenvalue, and the corresponding vibrational frequency must therefore be imaginary. The particular direction corresponding to a negative eigenvalue of the second derivative matrix is termed the *transition vector*. These saddle points are very important in the study of chemical reactions because they are the lowest barriers connecting reactants and products. They are the transition states of chemical reactions. One of the most difficult tasks in the study of chemical reactions is the localization, along the path of potential surface, of the transition states.

Optimization of minima

The algorithms used in geometry optimization in the programmes used in this thesis are those of Schlegel [1.23] and Broyden-Fletcher-Goldfarb-Shanno (BFGS) [1.24].

The Schlegel algorithm is used in the ab initio MO methods. This is a modified conjugate gradient algorithm which evaluates and utilizes the gradients each time the energy is computed.
Both methods use the standard quasi-Newton formula:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k H_k \mathbf{g}_k \qquad (1.67)$$

where $\mathbf{g}_k$ is the gradient at $\mathbf{x}_k$, $\mathbf{x}_{k+1}$ is the improved geometry and H_k is an approximation of the inverse Hessian matrix, α_k is a constant.

Within the Schlegel's method, at each step of the optimization, the gradient and the displacement vector are tested to see if the optimization has converged. The root-mean-square (r.m.s.) gradient and the absolute value of the largest component of the gradient must fall below their respective threshold. The estimated displacement must pass a similar test on the r.m.s. and absolute value of the largest component. If these four conditions are satisfied, the optimization is considered complete.

Within the BFGS method, as implemented in MOPAC-V5.0 [1.25] the geometry optimization is terminated when one or more of the following calculated quantity fall below their respective thresholds: 1) predicted change in geometry, 2) predicted change in energy, 3) current gradient norm. The norm σ of the gradient g is defined as:

$$\sigma = \sum g_i^2 \qquad (1.68)$$

The users can modified the threshold of the various optimization criteria, increasing the precision or switching-off some of them.

Location of transition states ,

The first step to locate a saddle point is to determine the reaction pathway. In the frequently used "reaction coordinate" approach [1.26], the path is constructed by successively incrementing a selected internal coordinate (e.g. bond length or angle) between its path limiting values. The remaining degrees of freedom are reoptimized at each step. The choice of reaction coordinate is not always obvious; a given choice of reaction coordinate can produce different pathways, depending upon whether one starts from "reactants" or from "products".
Other approaches, such as the Linear Synchronous-Transit (LST) [1.27], appear to avoid these and other difficulties of the "reaction coordinate" approach. In these methods the whole change in geometry is taken as the "reaction coordinate". It automatically generates a continuous path between the specified reactant and the product. Sometimes additional computational efforts are required to locate a proper transition state, but important information concerning the shape of the energy surface is obtained.

The resulting geometry can be refined using an eigenvalue-following (EF) optimization. The EF optimization procedure uses a Baker [1.28-1.30] implementation of Simon's method [1.31].

a) Eigenvalue-following (EF) method

For a transition state search, if we are in a region of the energy surface where the Hessian matrix has the required one negative eigenvalue, the usual Newton-Raphson (NR) step to get from the current point x_0 to a stationary point is a good step since attemps to minimize E along the modes with positive hessian eigenvalues and maximize E along the mode with negative eigenvalue.

On the other hand, if we are not in such a region of the energy surface, we have to get out of the current region to get into a region where the Hessian does have the correct structure. In this case the NR step is not appropriate; in the EF method, a shift parameter λ is added to the NR step. This allows one to go in a new region of the energy surface. In this new region the local surface characteristic have to be determined, this involves updating the Hessian.

The great advantage of this method is that it is capable to locate transition states even if the search is started in the wrong region of the energy surface.

1.17) THE MOLECULAR ELECTROSTATIC POTENTIAL (MEP)

The electric potential at a point r in vicinity of a given molecule is the electric force acting on a unit positive charge at that point caused by the nuclei and the electrons of the molecule [1.32].

The total electrostatic potential at point r is given by the difference between two terms:

V(r)= nuclear term- electronic term

Suppose that the molecule has M nuclei and N electrons then according to the Coulomb' s law the nuclear term $V_{nucl}(r)$ arising from the nuclei at a point r is:

$$V_{nucl}(r) = \sum_{A=1}^{M} \frac{Z_A}{|r - R_A|} \tag{1.69}$$

Z_A = charge of nucleus A

R_A = coordinates of nucleus A

To obtain the electric potential $V_{elec}(r)$ arising from the electrons we must integrate over the electronic coordinates. Let the electric charge associated to an electron i in the elementary volume dr^1 be given by $\rho(r^1)dr^1$ then the electric potential $V_{elec}(r)$ arising from the electron is:

$$V_{elec}(r) = \int (\rho(r^1) / |r - r^1|) dr^1 \tag{1.70}$$

where dr^1 goes over all space.

In terms of the density matrix the total electric potential is:

$$V(r)=\sum_{A=1}^{M}\frac{Z_A}{r-R_A}-\sum_{\mu=1}^{N}\sum_{\nu=1}^{N}P'_{\mu\nu}\int\frac{\phi\mu^*\phi\nu}{|r-r'|} \qquad (1.71)$$

The main use of the MEP is in obtaining a reliable indication of molecular interactions such as those between drugs and receptors, enzymes and substrate. The MEP has been used to study reactive sites of the molecules, hydrogen bonding etc.. There are many examples in the literature where the MEP has been used in the attempt to correlate biological activity with the electrostatic features, including nucleic acid bases [1.33, 1.34], enzyme active sites [1.35], the opiate receptor [1.36], neuroleptic drugs [1.37] , and DNA [1.38,1.39].

One area where the MEP is most informative is the study of the interactions between solute and solvent as described in the next section.

A major drawback in the use of the MEP in biological applications is that its calculation is rather time consuming. The length of the time taken to calculate $V(r)$ depends upon the number of two and three-centre nuclear-electron attraction integrals which must be calculated at each point. For a wavefunction described by n basis functions there are typically $n(n + 1)/2$ such integrals. Therefore for molecules over 10 atoms and larger basis set such as 6-31G** the time involved to calculate

the MEP is prohibititve unless huge computer resources are available.

1.18) SOLVATION

In theoretical studies of molecular systems and chemical reactions of biological interest, in many cases it is necessary to consider the effect of the solvent in the process. In fact, the properties and behaviour of biomolecules in their environment are considerabily influenced by the surrounding solvent. To date several strategies have been employed to investigate the effects of a solvent on these molecules and on the biochemical reactions in which they are involved. One of these strategies was developed by the group in Pisa [1.40-1.42]. In this classical continuum approach, the solvent is represented as a medium of structureless homogeneous polarizable dielectric interacting with a solute molecule, M, placed in a cavity.

The polarized dielectric is considered as a perturbation to the solute Hamiltonian, $\mathcal{H}^0$, which leads to a new Schrödinger equation:

$$(\mathcal{H}^0_M + V_\sigma)\,\Psi'_M = E'_M \Psi'_M \qquad (1.72)$$

where V_σ is the interaction operator

V_σ is defined in a semiclassical approximation i.e. limiting the solute-solvent interaction to the classical coulombic and

polarization terms, and representing the averaged distribution of the solvent as a continuum.

Basic outline of the method

The procedure basically relies on the calculation of the dielectric polarization charge distribution $\sigma(s)$ on the cavity surface related to the solvent reaction field. This charge distribution is represented by a number of discrete point charges q_i placed at the center of cavity surface elements (tesserae) ΔS_i of appropriate area. The value of q_i is given by:

$$q_i = \sigma(s_i)\, \Delta S_i \qquad (1.73)$$

where

$$\sigma(s) = [(\varepsilon-1)/4\pi\varepsilon]\, E(s) \qquad (1.74)$$

ε = dielectric constant of the solvent

$E(s)$ = electric field due, in first approximation, to the solute charge distribution of the unperturbated solute $\rho^0(r)$.

The electric field is modified by the mutual polarization of surface charges by a iterative procedure, until a final set of surface charges q^{0f} is obtained.

The next step consists in the introduction in $\mathcal{H}^0_M$ of the perturbative term:

$$V_\sigma = \Sigma_i\, q_i^{0f} /[r-r_i] \qquad (1.75)$$

r= solute coord. ; r_i= surface points coord.

Hence a new solute charge distribution $\rho^1(r)$ is obtained.

Further readjustments of $\sigma(s)$ and $\rho(r)$ are obtained by repeating the above procedure.

When the final solution of equation (1.72) is reached the information encoded in V_σ, Ψ'_M and E'_M can be further elaborated to get a number of different properties, among which the free energy, G, of the system:

$$G = E'_{tot} - 1/2 \int \rho'(r) V'_\sigma \, dr \qquad (1.76)$$

This quantity is directly comparable with E^0_{tot} because the latter, obtained in the frozen nuclei approximation, may be interpreted either as internal energy or as an electrostatic free energy. For this reason $E^0_{tot} = G^0_{tot}$ using non interacting electrons and nuclei as reference.

The electrostatic contribution to the solution free energy change ΔG_{sol}, at a given geometry of M is given by

$$\Delta G_{sol} = G - E^0_{tot} \qquad (1.77)$$

E'_{tot} and E^0_{tot} are the total energies computed at a given geometry of M using respectively the solution of equation (1.72) and the solution of the corresponding Schrödinger equation involving $\mathcal{H}^0_M$ only.

The solvent contribution to the nuclear repulsion energy is given by:

$$\sum_{A=1}^{M}\sum_{i=1}^{NTS}\frac{Z_A q_i}{R_{Ai}} \qquad (1.78)$$

where:

A = indices for nuclei A=1,2.....M

i = indices for tesserae i=1,2.....NTS

Z_A = atomic number of nucleus A

q_i = charge on tesserae i

R_{Ai} = distance between nucleus A and point charge on tesserae i

Compensation of surface charges

The sum of the surface charges $Q = \Sigma q_i$ should be equal and opposite in sign to the total charge of the solute, which is zero for neutral molecules. Within this method this is not the case, in fact in cavities of reasonable size, there is always a portion of $\rho(r)$ lying outside this boundary. This produces a systematic error which is corrected by the introduction of a compensation factor which distributes a charge -Q over the surface elements according to the sign and magnitude they have.

Description of the cavity

The cavity is represented as a system of interlocking spheres, centered on appropriate nuclei of the solute with radii R_1, R_2.....R_N (usually Van Der Waal's radii). The surface of the spheres is divided in "tesserae" created by a solid, the pentakisdodecahedron $\{ 3,5+\}_{1,1}$ in geodesic symbolism, inscribed

in the reference sphere. Elegant strategies have been introduced to describe σ at the interlocking spheres and to eliminate regions of the outer space too small for accomodating solvent molecules.

1.19) IMPLEMENTATION OF THE SOLVATION CONTINUUM MODEL IN THE GAUSSIAN 88 PROGRAM

One of the most difficult and time consuming aspects of this research project has been the implementation of the solvation continuum model in the Gaussian 88 (G88) programme [1.39].

The G88 program is organized in links grouped in overlays. Each overlay deals with different sections of the method, such as integral evaluation, SCF etc..

Once a route is generated by the program following the instruction given by the user, a routine: Chain calls the pertinent routines in the order given by the route. Some of the constant values are transferred through common blocks, other values are passed between the different parts of the program through Read&Write files which store data and matrices to be updated.

Overlay L100 reads and parses the route card, L200 reads the geometry, L300 sets up the basis set and computes one and two electron integrals, L400 makes the initial guess to the solution of the Roothaan equation, L500 solves the Roothaan equation, L600 does the population analysis, etc..

The implementation of the solvation in G88 consists in the modification of pre-existing links and in the addition of the new links L604 and L605.

Link L604 deals with the geometry of the cavity. It reads from the input the number of the spheres, their rays and their centers and calculates N points on the surface and N points on the normal to the cavity at a very small distance from the surface. The potentials in these points is computed in the link L605 as the sum of the contribution coming from all the overlaps of primitive gaussian functions multiplied by the fractional electron population. The next step is to calculate iteractively the dielectric polarization charge distribution on the cavity surface due to the solute-solvent interaction. The dielectric constant of the solvent is given in the input. The electric field on the surface cavity is firstly calculated as the derivative of the potential on the points on the cavity and then modified by the mutual polarization of surface charges. The calculations, all allocated in L605, proceed until a set of stationary points is reached, four iterations are default.

Links L301, L302, and L501 are modified. In L301 the solvent charge contribution to the nuclear potential energy is calculated. Matrix elements of the charge surface potential operator are computed in link L302, and added to the Hamiltonian of the solute in link L501; here the total and the free energies of the system are computed. The whole calculation is repeated until the Hamiltonian does not change. Three or four cycles are suggested. For more details see chapter 5.

1.20) References

1.1) M. Born and J.R. Oppenheimer; Ann. Physik, **84,** 457 (1927).

1.2) A. Szabo and N.S. Ostlund; "Modern Quantum Chemistry", Edition McMillan, New York, (1982).

1.3) W. Pauli, Z. Physik, **31**, 765 (1925).

1.4) C.C.J. Roothaan; Rev. Mod. Phys., **23,** 69 (1951).

1.5) G.G. Hall; Proc. Roy. Soc. (London) 10 **A205** 541 (1951).

1.6) J.A. Pople and R.K. Nesbet ; J. Chem. Phys. **22**, 571 (1954).

1.7) M.J.S. Dewar, J.A. Hashmall and C.G. Venier; J. Am. Chem. Soc. **90:8**, 1953 (1968).

1.8) T.H. Dunning, jr., P.J. Hay ; in Modern Theoretical Chemistry 3, Ed. H.F. Schaefer III , Plenum Press, New York; p. 1-27 (1977).

1.9) S.F. Boys; Proc. R. Soc. (London), Ser.A.**200,** 542 (1950).

1.10) S. Huzinaga "Gaussian Basis Sets for Molecular Calculations" Elsevier (1984).

1.11) W.J. Hehre, R.F. Stewart, J.A. Pople; J. Chem. Phys.; **51,** 2657 (1969).

1.12) J.S. Binkley, J.A. Pople, W.J. Hehre; J. Amer. Chem. Soc.; **102,** 939 (1980).

1.13) W.J. Hehre, R. Ditchfield, J.A. Pople; J. Chem. Phys.; **56,** 2257 (1969).

1.14) P.C. Hariharan, J. A. Pople; Theor. Chim. Acta: **28,** 213 (1973).

1.15) M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart; J. Am. Chem. Soc. **107** 3902-3909 (1985).

1.16) J.J.P. Stewart ; J. Comp. Chem.,**10** 209-220 (1989).

1.17) J.J.P. Stewart ; J. Comp. Chem.,**10** 221-264 (1989).

1.18) J.A. Pople, D.P. Santry and G.A. Segal ; J. Chem. Phys., **43** 129 (1965).

1.19) J.A. Pople, D.L. Beveridge "Approximate Molecular Orbital Theory" McGraw-Hill, New York, N.Y. (1970).

1.20) L. Oleari, L. DiSipio and G. De Michelis; Mol. Phys. **10** 97 (1966)

1.21) M.J.S. Dewar, W. Thiel; J. Am. Chem Soc., **99** 4899-4907 (1977).

1.22) T. Koopmans, Physica, **1**, 104 (1934)

1.23) H.B. Schlegel , J. Comp. Chem., **3(2)**, 214 (1982)

1.24) D.F. Shanno ; J. Optim. Theory Appl.; **46**, 87 (1985)

1.25) MOPAC-V5.0; J.J.J. Stewart; (private communication)

1.26) K. Jug , Theor. Chim. Acta, **54**, 263 (1980);

1.27) T.A. Halgren and W.N. Lipscomb; Chem. Phys. Lett.; **49(2)** 225 (1977)

1.28) J. Baker; J. Comp. Chem., 7(4) 385 (1986);

1.29) J. Baker; J. Comp. Chem., **8(5)** 1987 563, (1987);

1.30) J. Baker; J. Comp. Chem.,**4(3)**, 294. (1983);

1.31) J. Simons, P. Jorgenson, H. Taylor, J. Ozment ; J. Phys. Chem., **87**, 2745, (1983).

1.32) P.Politzer, D.G. Truhlar; "Chemical Applications of Atomic and Molecular Electrostatic Potential", Plenum Press, New York, (1981).

1.33) R. Bonaccorsi, E. Scrocco, J. Tomasi and A. Pullman; Theoret. Chim. Acta, **24**, 51, (1972);

1.34) D. Perahia and A. Pulman; Theoret. Chim. Acta, **48**, 263, (1978);

1.35) L.C. Allen in "Quantum Chemistry in Biomedical Sciences", (H. Weinstein and J.P. Green eds.), New York Academy of Sciences, New York, 383 (1981);

1.36) W.G. Richards "Quantum Pharmacology (2nd Edition), Butterworths, London (1983);

1.37) C. Petrongolo, H.J.T. Preston and J.J. Kaufman, Int. J. Quant. Chem., Vol. XIII, 457, (1978);

1.38) J.M. Burridge, P. Quarendon, C.A. Reynolds and P.J. Goodford, J. Mol. Graph., 5(3), 165 (1987);

1.39) R. Lavery and A. Pullman, in "New Horizons of Quantum Chemistry", (P.O. Lowdin and B. Pullman eds.), D. Reidel Publishing Company, 439, (1983);

1.40) S. Miertus, E. Scrocco and J. Tomasi ; Chem. Phys. **55** 117 (1981).

1.41) S. Miertus and J. Tomasi ; Chem. Phys. **65** 239 (1982).

1.42) J.L. Pascual-Ahuir, E. Silla, J. Tomasi and R. Bonaccorsi; J. Comp. Chem. **8**, 778 (1987).

1.43) GAUSSIAN 88; M.J. Frish , M. Head-Gordon, H.B. Schlegel, K. Raghavachari, J.S. Binkley, C. Gonzales, D.J. Defrees, D.J. Fox, R.A. Whiteside, R. Seeger, C.F. Melius , J. Baker, R. Martin, R.L. Kahn, J.J.P. Stewart, E.M. Fluder, S. Topiol, and J.A. Pople; Gaussian, Inc., Pittsburgh, PA; (1988).

CHAPTER 2

COMPARISON BETWEEN CALCULATED AND EXPERIMENTAL PROPERTIES OF POLYCYCLIC AROMATIC COMPOUNDS AND OTHER LARGE MOLECULES

INTRODUCTION

The heats of formation of the ground state of molecules of importance in biology must be known if we are to understand fully the energetic details of many biological processes. Biologically important molecular systems are, in most cases, large and contain often tens of atoms. Unfortunately, for the majority of molecules of this size, the available thermochemical data are very limited. Semiempirical methods, however, have now been developed to the point where the computation of theoretical molecular structures and the molecular potential energy surfaces can be carried out with high accuracy also for larger molecules.

The recent methods, AM1 [2.1] and PM3 [2.2, 2.3] derived from NDDO [2.4, 2.5] have been parametrized in order to reproduce, among other properties, experimental heats of formation and ionization potentials.

In this part of the thesis, a comparison between experimental and calculated heats of formation and ionization potentials for cyclic molecules having between 9 and 36 atoms is reported in order to assess the quality of these semiempirical methods in the study of biological systems. It is worth noting that the development of these methods is currently a rather active research area, and a critical assessment of the current state of this is still premature.

2.1) POLYCYCLIC AROMATIC COMPOUNDS

In this part of the thesis a wide range of different compounds have been studied, among them the polycyclic aromatic compounds. Accordingly, we have used this class of molecules as suitable test cases for the evaluation of the accuracy of semi-empirical SCF methods in calculating ΔH_f^0.

Among many wide classes of carcinogenic compounds are some polycyclic aromatic compounds [2.6]. Some of them are very powerful carcinogens, other require metabolic activation; hence they are often called "procarcinogens". The reactive species are electrophiles that react with nucleophilic sites of DNA by forming covalent bonds. In Figure 2.1 the metabolic activation of Benza[a]pyrene and its interaction with one of the bases of DNA is shown. The major adduct of Benza[a]pyrene is from the 10-position of the hydrocarbon to the 2-position of Guanine. Furthermore, the methylation in one specific region of the molecule can lead to different behaviour and carcinogenic power. Several relatively weak carcinogens (e.g. benz[a]anthracene) can be enzymatically methylated to give much more potent carcinogens (e.g., 7,12-dimethylbenz[a]anthracene). Substitution of a methyl group at carbon 11 of benzo[a]pyrene increases its tumorigenicity, whereas substitution at carbons 6-10 decreases its tumorigenicity. Among the benz[a]anthracenes, the 7- and 12-methyl isomers are the most reactive (Figure 2.2).

2.2) HEATS OF FORMATION

2.2.1) Methods

The standard enthalpy change ΔH_f^0 of 40 compounds in the gaseous phase at 298K has been computed by AM1 and the more recent PM3 methods (some of the data have been already published [2.7]). All the molecules examined here are ring compounds containing hydrogen, carbon, oxygen, and nitrogen atoms.

In the first stage of the calculations the geometry was optimized using the PRECISE key-word as implemented in the Mopac-V5.0 package [2.8], by which the optimization criteria are increased by a factor of 100. No geometrical assumptions were made. In a second stage, the norm of the gradient (i.e. the scalar of the vector of derivatives of the energy with respect to coordinates) was minimized by the use of the GNORM option. The value chosen for the gradient minimization was 0.1 Kcal/mol^{-1}Å^{-1}, for lower values no significant changes in ΔH_f^0 have been obtained. The last stage was to check if the resulting geometry was a minimum on the potential energy surface. In this case, a force calculation was run, and the Hessian matrix was calculated. Its eigenvalues were found all positive.

2.2.2) Results

The heats of formation of molecules as calculated by PM3 and AM1 are given in Table 2.1. Here the calculated values of heat of formation in Kcal/mol are compared with experimental values

[2.9, 2.10]. In Figure 2.3, the experimental values are plotted versus the predicted heats of formation calculated by PM3 and AM1, respectively. With PM3 and AM1 the average difference is 7.42 and 11.9 Kcal/mol respectively. The majority of the differences are positive.

Some of the errors of AM1 have been reduced by PM3. As far as the Furanoids are concerned, the significant differences between the experimental heats of formation and those calculated by AM1, have been corrected by PM3. In Table 2.1, we can see that the computed heats of formation for Furan calculated by PM3 and AM1 differ from the experimental values of 4.3 and 11.3 Kcal/mol respectively. The lowering of the heat of formation by PM3 with respect to AM1 is 7.0 Kcal/mol. The same behaviour can be noted for Dibenzofuran, Furfural and Furandione where the difference between the two methods are of 10.8, 11.3 and 13.7 Kcal/mol respectively. In table 2.2 the geometries of Furan and Furandione calculated by PM3 and AM1 are reported. The bond lengths are in Angstrom the bond angles in degrees. We can see that no significant changes in geometry are found between the structures computed by PM3 and AM1.

When polycyclic aromatic linear, angular and condensed compounds (see Figure 2.4 and 2.5) are considered, although the errors in computed heats of formation have been drastically reduced by PM3, there are still few molecules in which the difference between experimental and computed values is > 10.0 Kcal/mol.

In tables 2.3 and 2.4 the experimental bond lengths in Angstrom for Naphthalene and Anthracene have been compared to PM3 and AM1 values. The two semiempirical methods give the same results and when compared with the X-ray values, only a slight tendency to stretch the bonds can be observed.

The linear polycyclic compounds become increasingly reactive as additional rings are added, furthermore the energy difference between the highest occupied molecular orbital HOMO and the lowest unoccupied molecular orbital LUMO decreases as rings are added. In table 2.5 the HOMO and LUMO energy values in eV calculated by AM1 and PM3 for linear, angular and condensed polycyclic aromatic compounds are reported. As far as linear compounds are concerned the data are in agreement with the above mentioned experimental observation.

Many force constants for torsional bending are very low compared to bond length stretching. This implies that changes in torsional angle can result from very small changes in energy. In table 2.6 the internal coordinates for Phenanthrenequinone are reported. The calculations were carried out with PM3 method. The geometries were optimized with Broyden-Fletcher-Goldfarb-Shanno (BFGS) [2.11] and Davidon-Fletcher-Powell (DFP) [2.12] methods. The BFGS optimization method is the default one in the more recent version of Mopac (Mopac 5.0), and DFP was the default optimization method in the former versions of Mopac. The PRECISE and GNORM=0.1 options were used in both calculations. The optimization with DFP gives a planar geometry and a heat of

formation of -10.46 Kcal/mol. The optimization with BFGS gives a distorted geometry and -11.72 Kcal/mol as heat of formation. A relatively small difference in heat of formation of 1.26 Kcal/mol yields quite different torsional angles.

2.3) IONIZATION POTENTIALS

The ionization potentials of 29 compounds have been computed by AM1 and PM3. The value was taken, according to Koopman's theorem, as the negative of the HOMO energy. Full optimization of each geometry was performed in every calculations. The ionization potential of the molecules as calculated by each method are given in table 2.7. Here the calculated values of the ionization potentials are compared with the experimental values [2.13], the differences between calculated and experimental values are also reported. In Figure 2.6 the experimental values were plotted versus the predicted ionization potentials calculated by PM3 and AM1, respectively. As we can see the results are quite reasonable with both methods. Both Hamiltonians give always HOMO energies higher than experimental ionization potentials. The average differences between experimental and theoretical data are 0.74 eV for PM3 and 0.67 eV for AM1.

2.4) CONCLUSIONS

Considerations on the performance of these methods for calculations of heats of formation and ionization potentials, as reported herein, suggest that although PM3 represents a remarkable improvement on AM1 in calculating heats of formations , AM1 performs slightly better than PM3 in calculating ionization potentials.

Figure 2.1 METABOLIC ACTIVATION OF BENZOPYRENE

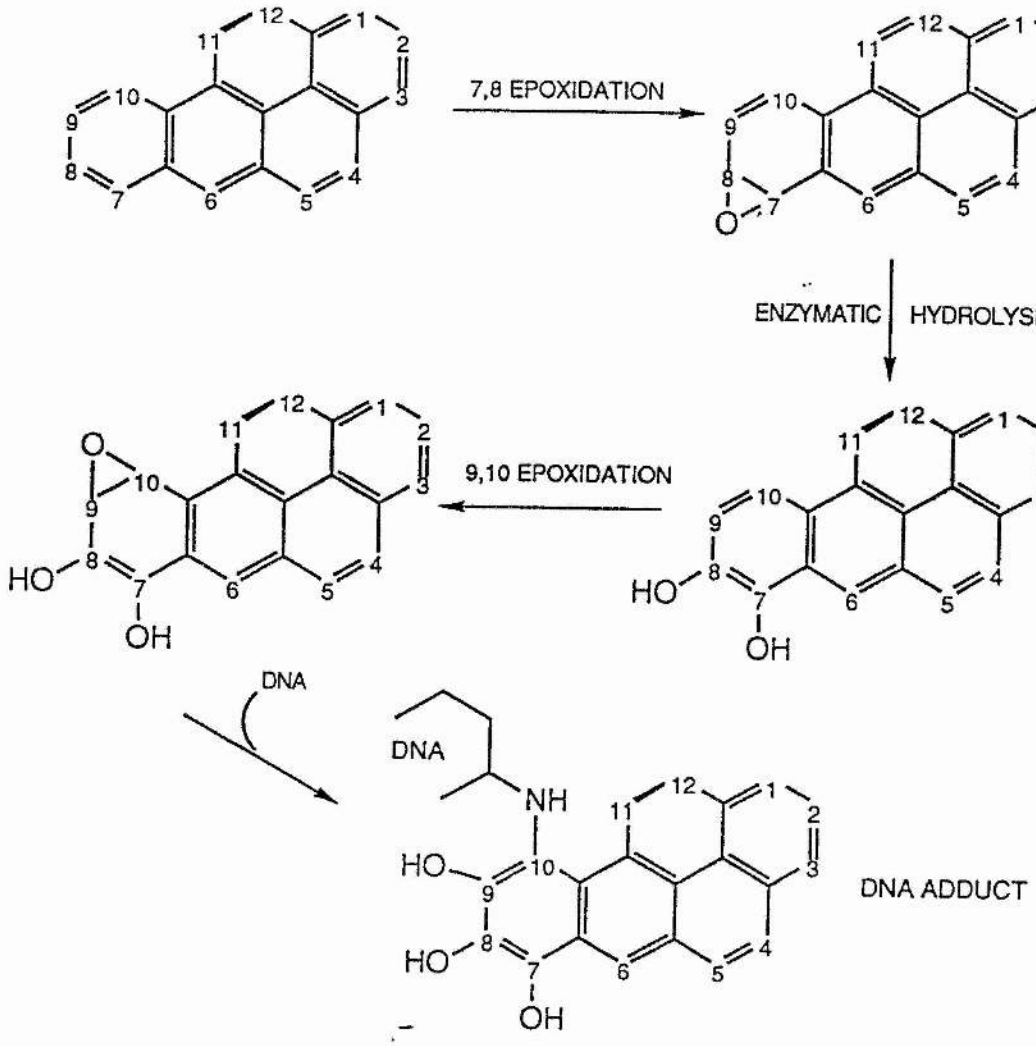
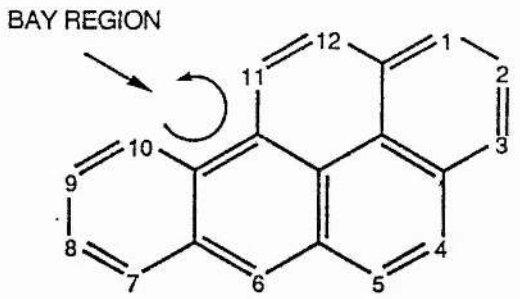


Figure 2.2

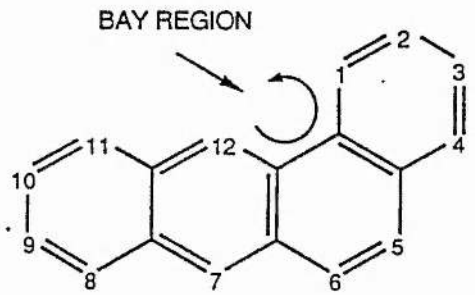
BENZO[a]PYRENE



Enhancement of tumorigenicity: 11-Methyl;

Reduction of tumorigenicity: 6,10-diMethyl;

BENZA[a]ANTHRACENE



Enhancement of tumorigenicity: 12-Methyl; 7,12-diMethyl;

Reduction of tumorigenicity: 5,7-diMethyl; 5,7,12-triMethyl;

Table 2.1

Empirical formula	Chemical name	Heat of formation Kcal/mol			Difference[a]		
		Exp.	PM3	AM1	PM3	AM1	ref
$C_{10}H_8$	NAPHTHALENE	35.9	40.7	40.6	4.8	4.7	2.9
$C_{11}H_{10}$	1-METHYL NAPHTHALENE	27.9	32.6	33.9	4.7	6.0	2.10
$C_{14}H_{10}$	PHENANTHRENE	49.6	55.0	57.4	5.4	7.8	2.9
$C_{14}H_{10}$	ANTHRACENE	55.2	61.7	62.9	6.5	7.7	2.9
$C_{16}H_{10}$	PYRENE	53.9	64.1	67.3	10.2	13.4	2.10
$C_{18}H_{12}$	1,2 BENZANTHRACENE	70.0	74.5	78.2	4.5	8.2	2.9
$C_{18}H_{12}$	NAPHTHACENE	69.9	84.3	86.9	14.7	17.3	2.9
$C_{18}H_{12}$	CHRYSENE	64.5	70.8	76.2	6.3	11.7	2.9
$C_{20}H_{12}$	2,3 BENZOPYRENE	71.9	81.7	87.5	9.8	15.6	2.10
$C_{22}H_{14}$	PENTACENE	89.5	107.9	111.8	18.4	22.3	2.10
$C_{22}H_{14}$	1,2,3,4 DIBENZO ANTHRACENE	84.6	86.9	95.6	2.3	11.0	2.10
C_4H_4O	FURAN	-8.3	-4.0	3.0	4.3	11.3	2.9
$C_5H_{10}O$	TETRAHYDRO-2H-PYRAN	-53.4	-57.4	-67.1	-4.0	-13.7	2.9
$C_6H_{10}O$	CYCLOHEXANONE	-54.0	-60.2	-63.3	-6.2	-9.3	2.9
$C_{10}H_8O$	1-NAPHTHOL	-7.1	-4.1	-3.0	3.0	4.1	2.9

continue

Table 2.1

Empirical formula	Chemical name	Heat of formation Kcal/mol			Difference		
		Exp.	PM3	AM1	PM3	AM1	ref
$C_{10}H_8O$	2-NAPHTHOL	-7.2	-4.6	-3.8	2.6	3.4	2.9
$C_{12}H_8O$	DIBENZOFURAN	19.9	25.5	36.3	5.6	16.4	2.9
$C_5H_4O_2$	FURFURAL	-36.1	-37.3	-26.0	-1.2	10.1	2.9
$C_6H_4O_2$	BENZOQUINONE	-29.4	-31.5	-25.1	-2.1	4.3	2.9
$C_6H_6O_2$	HYDROQUINONE	-63.4	-66.1	-65.7	-2.7	-2.3	2.9
$C_6H_{12}O_2$	2-METHOXYTETRAHYDROPYRAN	-95.5	-95.9	-111.7	-0.4	-16.2	2.9
$C_{10}H_6O_2$	1,4 NaPHTHAQUINONE	-26.3	-22.8	-15.9	3.5	10.4	2.9
$C_{11}H_8O_2$	α–NAPHTHOIC ACID	-53.3	-46.7	-47.2	6.6	6.1	2.9
$C_{11}H_8O_2$	β–NAPHTHOIC ACID	-55.6	-48.9	-49.3	6.7	6.3	2.9
$C_{14}H_8O_2$	9,10 ANTHRAQUINONE	-22.8	-13.2	-6.0	9.6	16.8	2.9
$C_{14}H_8O_2$	9,10 PHENANTHRAQUINONE	-33.3	-11.7	-1.0	21.6	32.3	2.9
$C_{18}H_{10}O_2$	5,12 NAPHTHACENEQUINONE	-8.1	3.2	12.2	11.3	20.3	2.9
$C_{18}H_{10}O_2$	1,2 BENZANTHRACENE 9,10 DIONE	-35.6	8.7	16.5	44.3	52.1	2.9
$C_4H_2O_3$	2,5 FURANDIONE	-95.2	-90.1	-76.4	5.1	18.8	2.9
$C_5H_4O_3$	3-METHYL 2,5 FURANDIONE	-106.9	-100.7	-85.7	6.2	20.7	2.9

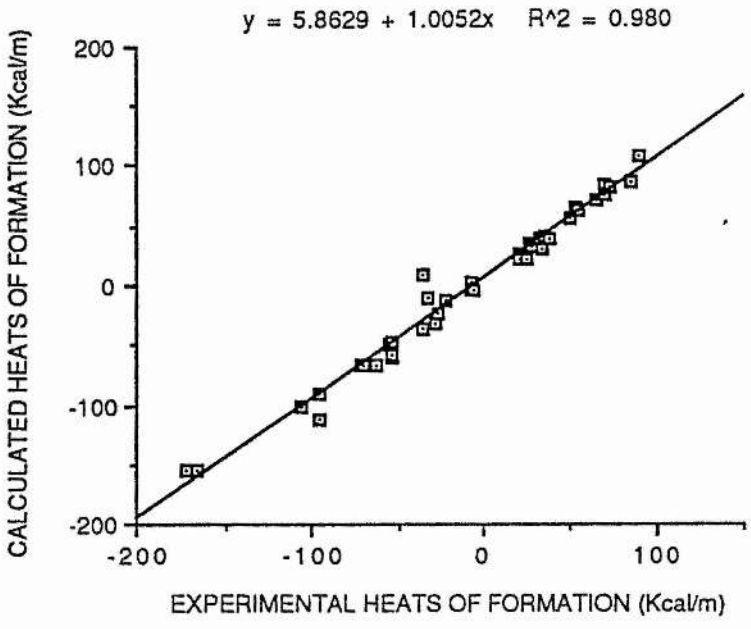
continue

Table 2.1

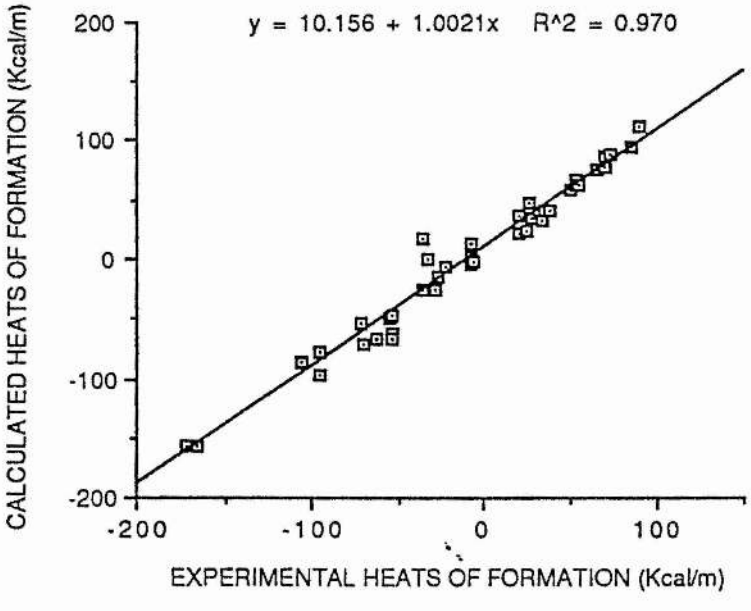
Empirical formula	Chemical name	Heat of formation Kcal/mol			Difference		
		Exp.	PM3	AM1	PM3	AM1	ref
$C_8H_6O_4$	ISOPHTHALIC ACID	-166.4	-155.1	-156.9	11.3	9.5	2.9
$C_8H_6O_4$	TEREPHTHALIC ACID	-171.6	-154.7	-156.2	16.9	15.4	2.9
C_5H_5N	PYRIDINE	33.6	30.4	32.0	-3.2	-1.6	2.9
C_6H_7N	4-METHYL PYRIDINE	24.9	20.8	24.2	-4.1	-0.7	2.9
C_7H_7N	ANILINE	20.8	21.3	20.5	0.5	-0.3	2.9
$C_{10}H_9N$	1-NAPHTHYLAMINE	37.7	39.6	40.3	1.9	2.6	2.9
$C_{10}H_9N$	2-NAPHTHYLAMINE	32.0	38.5	39.1	7.4	7.1	2.9
$C_7H_7O_2N$	2-AMINOBENZOIC ACID	-70.7	-67.2	-70.8	3.5	-0.1	2.9
$C_{10}H_7O_2N$	1-NITRONAPHTHALENE	26.6	35.4	46.9	8.8	20.3	2.9
$C_4H_4O_2N_2$	URACIL	-72.4	-67.7	-53.9	4.7	18.5	2.9

a) ΔH_f(exp)- ΔH_f(method)

Figure 2.3

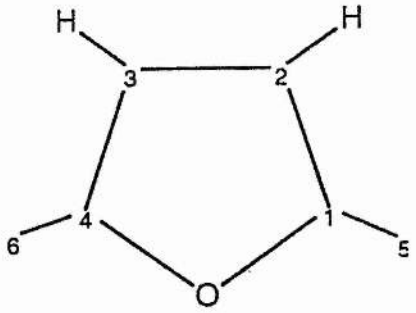


PM3



AM1

Table 2.2

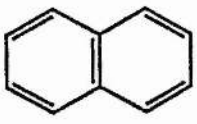


FURAN 5=H 6=H
FURANDIONE 5=O 6=O

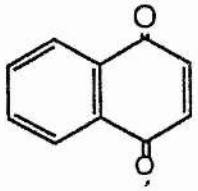
	FURAN		FURANDIONE	
Interatomic distances (Å)	PM3	AM1	PM3	AM1
O 1	1.378	1.400	1.400	1.409
1 2	1.373	1.380	1.503	1.497
2 3	1.441	1.448	1.342	1.349
3 4	1.373	1.380	1.503	1.497
4 O	1.378	1.400	1.400	1.409
4 6	1.085	1.086	1.200	1.217
C H	1.085	1.086	1.085	1.086
Bond Angles (degrees)				
4 O 1	106.9	106.6	106.9	107.6
O 1 2	110.2	110.1	109.0	108.3
1 2 3	106.3	106.6	107.6	108.0
2 3 4	106.3	106.6	107.6	108.0
3 4 O	110.2	110.1	109.0	108.3

FIGURE 2.4

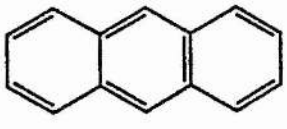
Linear Polycyclic Compounds and their corresponding Quinones



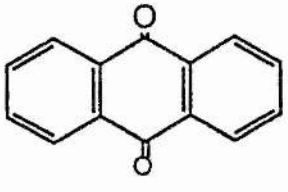
NAPHTHALENE



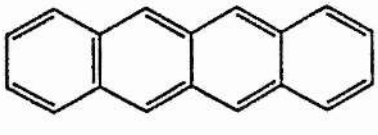
NAPHTHALENE QUINONE



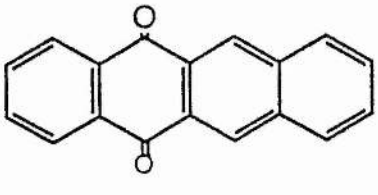
ANTHRACENE
noncarcinogenic



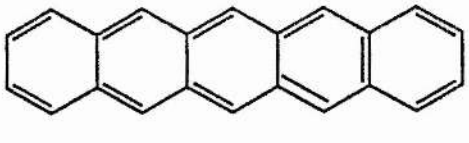
ANTHRACENEDIONE



NAPHTHACENE
noncarcinogenic

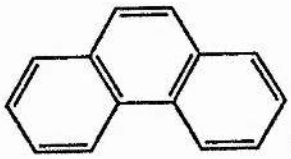


NAPHTHACENEDIONE

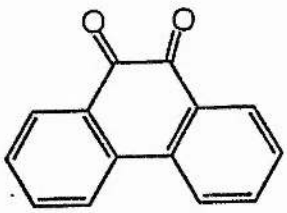


PENTACENE

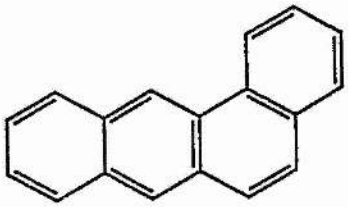
FIGURE 2.5
Angular and Condesed Polycyclic Compounds and their corresponding Quinones



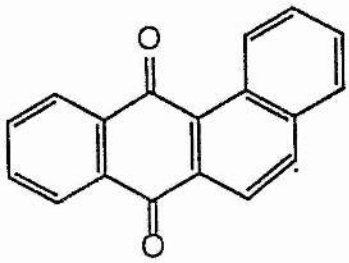
PHENANTHRENE



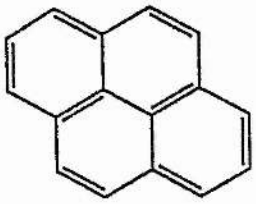
PHENANTHRENEQUINONE



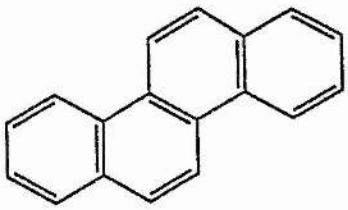
1,2 BENZANTHRACENE



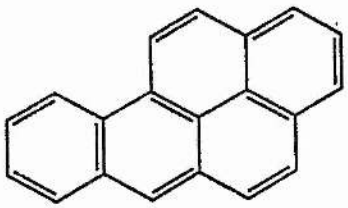
1,2 BENZANTHRACENEDIONE



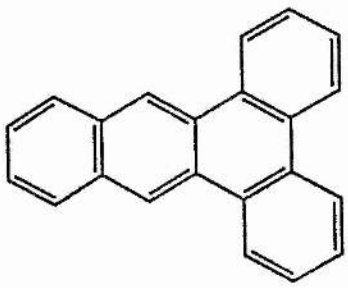
PYRENE



CHRYSENE



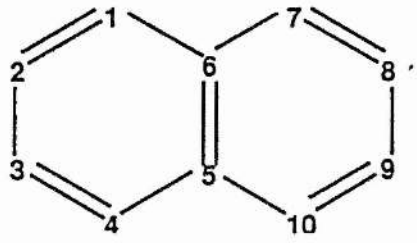
BENZOPYRENE



DIBENZOANTHRACENE

Table 2.3

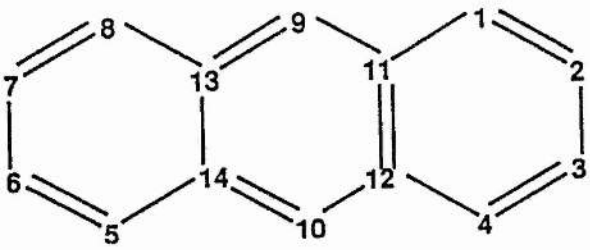
NAPHTHALENE



Bonds (Å)	X-RAY	PM3	AM1
1-2, 3-4, 7-8, 9-10	1.37	1.37	1.37
5-6	1.39	1.41	1.41
3-2, 8-9	1.40	1.42	1.42
1-6, 6-7, 4-5, 5-10	1.43	1.42	1.42

Table2.4

ANTHRACENE



BONDS (Å)	X-RAY	PM3	AM1
2-3, 7-6	1.39	1.43	1.43
1-2, 3-4, 5-6, 7-8	1.36	1.36	1.37
1-11, 4-12, 8-13, 5-14	1.42	1.43	1.43
9-11, 9-13, 10-12, 10-14	1.39	1.40	1.40
13-14, 11-12	1.44	1.42	1.43

Table 2.5

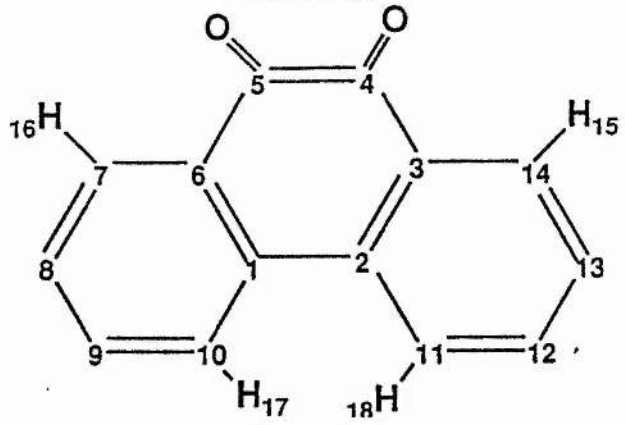
POLYCYCLIC COMPOUNDS

Linear compounds	number of rings	HOMO PM3	HOMO AM1	LUMO PM3	LUMO AM1	DIFFERENCE PM3	DIFFERENCE AM1
PENTACENE	5	-7.49	-7.61	-2.82	-2.93	4.67	4.68
NAPHTHACENE	4	-7.75	-7.78	-2.63	-2.75	5.12	5.12
ANTHRACENE	3	-8.12	-8.25	-2.24	-2.37	5.88	5.78
NAPHTHALENE	2	-8.71	-8.84	-1.67	-1.81	7.04	7.03

Angular and condensed compounds	number of rings	HOMO PM3	HOMO AM1	LUMO PM3	LUMO AM1	DIFFERENCE PM3	DIFFERENCE AM1
BENZOPYRENE	5	-7.92	-8.04	-2.51	-2.62	5.41	5.42
BENZANTHRACENE	4	-8.21	-8.33	-2.21	-2.33	6.00	6.00
CHRYSENE	4	-8.37	-8.50	-0.98	-1.08	7.39	7.42
PYRENE	4	-8.13	-8.25	-2.29	-2.41	5.84	5.84
PHENANTHRENE	3	-8.62	-8.74	-1.81	-1.93	6.81	6.81

Energies expressed in eV

Table 2.6



PHENANTHRENEQUINONE

	PM3/DFP	PM3/BFGS
BOND LENGTHS Å		
1,2	1.467	1.467
2,3 6,1	1.404	1.405
3,4 5,6	1.484	1.485
4,5	1.529	1.527
4,O 5,O	1.215	1.213
O,O	2.747	2.912
17,18	1.754	1.821
3,14 7,6	1.398	1.394
13,14 7,8	1.389	1.391
12,13 8,9	1.390	1.391
2,11 1,10	1.400	1.397
7,16 14,15	1.098	1.097
15,O 16,O	2.511	2.680
BOND ANGLES (degrees)		
1,2,3 6,1,2	121.0	120.4
2,3,4 5,6,1	121.1	119.4
3,4,5 4,5,6	117.8	114.4
3,4,O 6,5,O	122.1	123.5
4,5,O 5,4,O	120.1	122.1
4,3,14 7,6,5	119.1	120.5
DIHEDRALS (degrees)		
O,4,5,O	0.036	-42.45
10,1,2,11	0.054	-12.72

Table 2.7

Empirical formula	Chemical name	Ionization Potential eV			Difference[a]	
		Exp.	PM3	AM1	PM3	AM1
$C_{10}H_8$	NAPHTHALENE	8.25	8.84	8.71	0.59	0.46
$C_{11}H_{10}$	1-METHYL NAPHTHALENE	8.01	8.71	8.58	0.70	0.57
$C_{14}H_{10}$	ANTHRACENE	7.47	8.25	8.12	0.78	0.65
$C_{14}H_{10}$	PHENANTHRENE	7.86	8.74	8.62	0.88	0.76
$C_{16}H_{10}$	PYRENE	7.45	8.25	8.13	0.80	0.68
$C_{18}H_{12}$	1,2 BENZANTHRACENE	7.47	8.33	8.21	0.86	0.74
$C_{18}H_{12}$	NAPHTHACENE	7.04	7.87	7.75	0.83	0.71
$C_{18}H_{12}$	CHRYSENE	7.60	8.50	8.37	0.90	0.77
$C_{20}H_{12}$	3,4 BENZOPYRENE	7.39	8.04	7.92	0.65	0.53
$C_{22}H_{14}$	PENTACENE	6.74	7.61	7.49	0.87	0.75
C_4H_4O	FURAN	8.85	9.38	9.32	0.53	0.47
$C_5H_{10}O$	TETRAHYDRO-2H-PYRAN	9.46	10.57	10.35	1.11	0.89
$C_6H_{10}O$	CYCLOHEXANONE	9.14	10.48	10.30	1.34	1.16
$C_{10}H_8O$	1-NAPHTHOL	7.85	8.59	8.46	0.74	0.61
$C_{10}H_8O$	2-NAPHTHOL	7.90	8.72	8.64	0.82	0.74
$C_{12}H_8O$	DIBENZOFURAN	8.09	9.02	8.94	0.93	0.85
$C_5H_4O_2$	FURFARAL	9.50	9.73	9.73	0.23	0.23
$C_6H_4O_2$	BENZOQUINONE	9.96	10.92	10.88	0.96	0.92

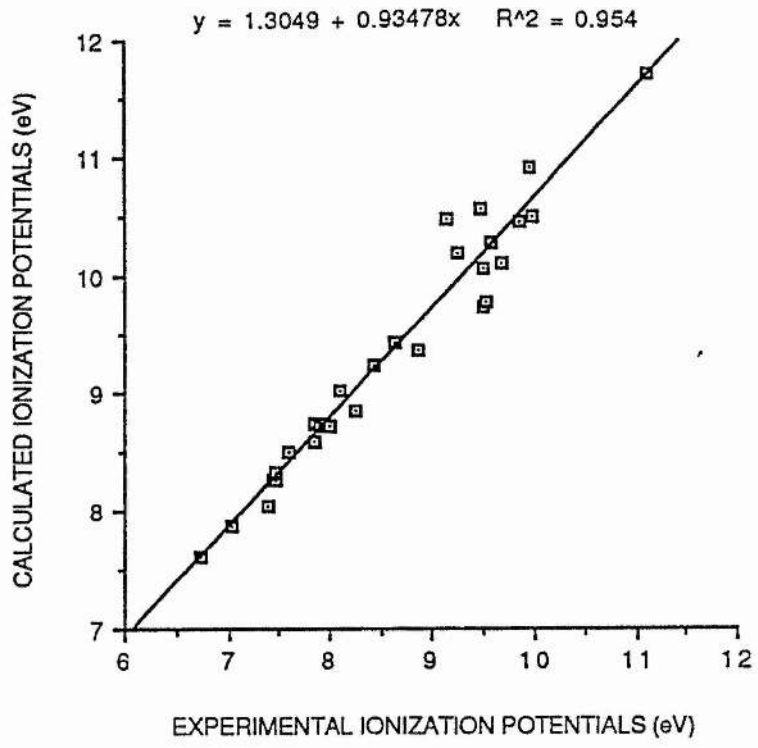
continue

Table 2.7

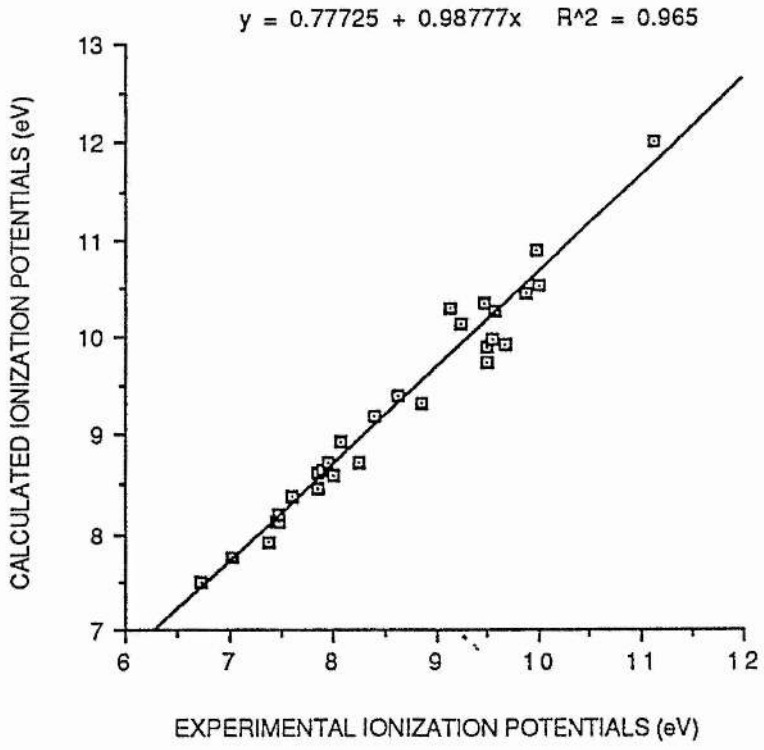
Empirical formula	Chemical name	Ionization Potential eV			Difference	
		Exp.	PM3	AM1	PM3	AM1
$C_6H_6O_2$	HYDROQUINONE	7.95	8.75	8.73	0.80	0.78
$C_{10}H_6O_2$	1,4 NAPHTHOQUINONE	9.56	10.29	10.26	0.73	0.70
$C_{13}H_8O_2$	XANTHONE	8.42	9.23	9.19	0.81	0.77
$C_{14}H_8O_2$	9,10 ANTHRAQUINONE	9.25	10.19	10.12	0.94	0.87
$C_{14}H_8O_2$	9,10 PHENANTHRAQUINONE	8.64	9.44	9.39	0.80	0.75
$C_4H_2O_3$	2,5 FURANDIONE	11.11	11.71	12.02	0.60	0.91
$C_8H_6O_4$	ISOPHTHALIC ACID	9.98	10.50	10.52	0.52	0.54
$C_8H_6O_4$	TERÉPHTHALIC ACID	9.86	10.45	10.44	0.59	0.58
C_5H_5N	PYRIDINE	9.66	10.10	9.93	0.44	0.27
C_6H_7N	4-METHYL PYRIDINE	9.50	10.06	9.89	0.56	0.39
$C_4H_4O_2N_2$	URACIL	9.53	9.79	9.97	0.26	0.44

a) eV(exp)- eV(method)

FIGURE 2.6



PM3



AM1

2.5) REFERENCES

2.1) M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart; J. Am. Chem. Soc., **107,** 3902 (1985);

2.2) J.J.P. Stewart ; J. Comp. Chem.,**10,** 209 (1989);

2.3) J.J.P. Stewart ; J. Comp. Chem., **10,** 221 (1989);

2.4) J.A. Pople, D.P. Santry and G.A. Segal ; J. Chem. Phys., **43** 129 (1965);

2.5) J.A. Pople, D.L. Beveridge; "Approximate Molecular Orbital Theory" McGraw-Hill, New York, N.Y. (1970);

2.6) P. Politzer, F.J. Martin; "Chemical Carcinogens" Elsevier (1988);

2.7) P. Scano and C. Thomson; J. Comp. Chem.,**12,** 172 (1991);

2.8) MOPAC 5, J.J.P. Stewart; (private communication);

2.9) J.B. Pedley, R.D. Naylor and S.P. Kirby; "Thermochemical data of Organic Compounds", London, Chapman and Hall, (1986);

2.10) TRC "Thermodynamic Tables Hydrocarbons", Thermodynamics Research Center, The Texas A & M University System, College Station, Texas;

2.11) D.F. Shanno; J. Optim. Theory Appl., **46** 87 (1985);

2.12) W.C. Davidon; Comp. J., **10** 406 (1968);

2.13) R.D. Levin and S.G. Lias; "Ionization Potentials and Appearance Potential Measurements", 1971-1981, Natl. Stand. Ref. Data Ser., Natl. Bur. Stnd. 71, (1982) Cat.No C13.48:71

CHAPTER 3

LIPID PEROXIDATION

INTRODUCTION

In this chapter, a complex biological mechanism such as Lipid Peroxidation is studied by means of semiempirical methods. The chapter is divided in two parts. The first one is a brief discussion on the biological aspects, and a description of the main features of the phenomenon taken into account in the calculations. The second part deals with the computational approach, with results and discussion.

CHAPTER 3.1: BIOLOGICAL BACKGROUND

Introduction

One of the most important characteristics of a eucaryotic cell is that it contains a cytoplasm separated from the environment by a membrane. The composition of cell membranes characterizes and determines the different morphology and functions of mammalian cells. In the electron microscope, the most prominent feature of mammalian cells is the abundance of subcellular organelles, each having its own peculiar membrane. All of these membranes conform to the same general pattern as plasma membranes. They have two layers of lipid molecules with a hydrophobic centre, and a number of protein molecules embedded in the membrane. The lipids provide the structural backbone with the proteins having a more functional role.

Almost all of the lipid molecules in biological membranes share one important feature- they are amphiphatic. That is, they have two different parts: one charged or polar which is stable in aqueous solution, the other a non-polar aliphatic hydrocarbon part which is more stable in non-aqueous environments. The most abundant lipids are phospholipids. The hydrophobic moiety of these lipids is derived from long-chain fatty acid molecules which are esterified, via a glycerol molecule and phosphate ester, with a hydrophilic base.

The most common long-chain fatty acids contain 14-24 carbon atoms. Some of them are unsaturated, and they have cis-double bonds separated by a methylene group so that conjugation does not occur.

COMMON FATTY ACID MOIETIES OCCURRING IN PHOSPHOLIPIDS

STRUCTURE	NOTATION	TRIVIAL NAME
$CH_3(CH_2)_{14}\ CO_2H$	16:0	Palmitic acid
$CH_3(CH_2)_{16}\ CO_2H$	18:0	Stearic acid
$CH_3(CH_2)_5CH{=}CH\ (CH_2)_7CO_2H$	16:1	Palmitoleic acid
$CH_3(CH_2)_7CH{=}CH\ (CH_2)_7CO_2H$	18:1	Oleic acid
$CH_3(CH_2)_4(CH{=}CH\ CH_2)_2(CH_2)_6CO_2H$	18:2	Linoleic acid
$CH_3CH_2\ (CH{=}CHCH_2)_3\ (CH_2)_6\ CO_2H$	18:3	Linolenic acid
$CH_3(CH_2)_4\ (CH{=}CHCH_2)_4\ (CH_2)_2\ CO_2H$	20:4	Arachidonic acid

The integrity of biological membranes plays a central role in controlling the physiological functionality of the cell organelles and of the cell itself. Any disturbance occurring in the composition

of membranes may give rise to pathological phenomena and in last instance to the death of the cell.

Fragmentation of the unsaturated fatty acid chains of phospholipid membrane components, due to oxidative stress, can cause disruption of plasma or subcellular membranes.

By the term "oxidative stress", we denote a cellular situation characterized by an elevation in the steady-state concentration of a variety of reactive oxygen species. Oxidative stress occurs if the balance between the cellular antioxidant defenses (superoxide dismutase, catalase, glutathione peroxidase, Vitamins E and C etc.) and the mechanism triggering oxidative conditions is impaired. The formation of reactive oxygen species is a consequence of a wide variety of physiological and pathological factors. Among them we may note the important examples of dietary deficencies of antioxidant compounds such as Vitamin C and Vitamin E.

3.1.1) Lipid Peroxidation

Lipid Peroxidation is a complex process that occurs in biological membranes which are rich in Polyunsaturated Fatty Acids (PUFAH).

A characteristic property of PUFAH is that they undergo oxidation in radical chain reactions [3.1]. An oxidising radical species (R•) can interact with a Polyunsaturated Fatty Acid (PUFAH) resulting in hydrogen atom abstraction and the formation of a Fatty Acid radical (PUFA•):

$$R\bullet + PUFAH \rightarrow RH + PUFA\bullet \qquad (3.1)$$

This reaction is the initiation step of the process. Reaction of PUFA• with molecular oxygen yields the corresponding peroxy radical (PUFAOO•) and chain propagation ensues, leading ultimately to degradation of the lipid to a wide range of biologically active products. Among them are lipid hydroperoxides, lipid alcohols, epoxy-fatty acids, aldehydes, and gases such as ethane and pentane. Some of these species have pronounced biological effects [3.2, 3.3]. Moreover, these products of a free radical reaction, which may have been initiated in some specific region, can diffuse considerable distances as compared with the cellular dimension [3.4].

The formation and diffusion of the Lipid Peroxidation secondary products can result in abnormal physiological processes and cell damage. In fact some of them were shown in extensive studies 1) to be strong inhibitors of protein synthesis [3.5] 2) to form DNA adducts [3.6], and 3) to be mutagenic [3.7-3.11]. On the other hand, it is also generally found that membranes of many neoplastic cells, especially liver hepatocytes have a much reduced ability to undergo stimulated Lipid Peroxidation than do normal cells [3.12-3.14]. The major reason of this occurence is the concentration of α-tocopherol (the most effective free radical scavanger) in the tumoral hepatocytes [3.15, 3.16].

It is interesting to note that an increase of antioxidant capacity and a decrease of capacity to undergo Lipid Peroxidation, found in

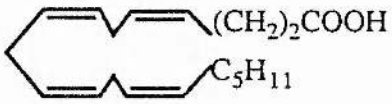
tumoral hepatocytes, was found also in normal dividing liver cells [3.17].

3.1.2) Chemistry of Lipid Peroxidation

Lipid Peroxidation is considered to be an important feature in the cellular injury which is due to oxidative stress brought about by chemical agents such as CCl_4, paraquat and ozone, and physical phenomena such as ionizing radiation. In these instances, Lipid Peroxidation is thought to be a consequence of the production of highly reactive free radicals that act as initiators of peroxidative chain reactions.

The PUFAH constituents of biological membranes (mainly Arachidonic and Linoleic acid) are characterized by the presence of a dienic methylene-interrupted system.

Linoleic acid



Arachidonic acid

The process of free radical autoxidation can cause a well defined disturbance in the structure of PUFAH. In fact, an electronic rearrangement of the double bonds, originally present in PUFAH, takes place [3.18] leading to a mixture of conjugated diene hydroperoxide products [3.19].

Although much work has been published on Lipid Peroxidation, details of the chemical mechanisms involved in the process are still unknown. While for many years the classical scheme proposed by Dahle [3.18] has been taken as the basis of the Lipid Peroxidation, in the last decade the hypothesis due to Porter's group [3.20, 3.21] has become widely accepted, and it is also the rationale at the molecular level, for this biological process.

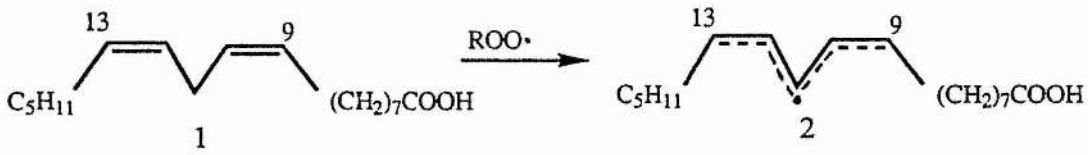
3.1.3) Porter scheme

The classical mechanism for autoxidation involves two propagation steps, a slow step (1) involving H-atom transfer and a second step (2) in which oxygen (generally in the singlet state [3.2, 3.4, 3.22, 3.23]) adds to the intermediate allylic radical to give a peroxide radical:

$$R\text{-}H + ROO\bullet \dashrightarrow R\bullet + ROOH \qquad (1)$$

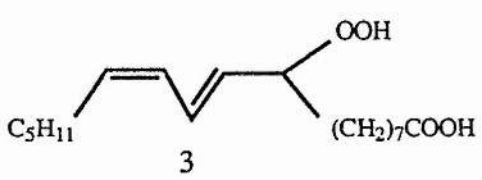
$$R\bullet + O_2 \dashrightarrow ROO\bullet \qquad (2)$$

The primary products formed from linoleate oxidation:



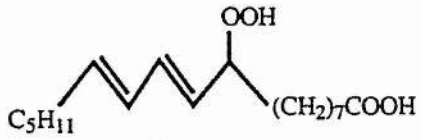
are hydroperoxides formed by oxygen addition at either the C9 or C13 atoms of the bis-allilyc radical 2. Four of these products have

been identified, and two of them (3,4) have trans,cis-conjugated diene stereochemistry while the other two (5,6) have trans,trans stereochemistry [3.20, 3.21].

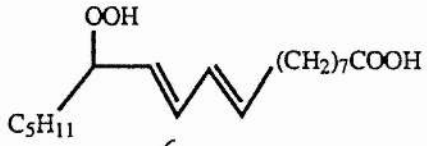


3

4



5



6

The hydroperoxides formed are thermally unstable and readily decompose into a host of secondary products.

The distribution of the primary products 3-6 depends on the conditions of autoxidation. It has been demonstrated that 1) the oxygen addition is equally favourable at both C9 and C13 of the intermediate bis-allilyc radical 2 [3.21], 2) higher autoxidation temperatures give rise to more trans,trans products 3) higher concentrations of linoleate and/or Vitamin E give more cis,trans products.

Porter's scheme is illustrated in Figure 3.1. The scheme shown applies to an initial oxygen addition at C9 of the pentadienyl radical 2, but it is understood that similar reaction occurs at C13. The following steps are involved:

1) Hydrogen abstraction from the bis-allylic carbon 11 of the acid 1 yields the W cis,cis radical 2 [3.21, 3.24].

2) Singlet oxygen reversibly reacts with 2 to give the trans,cis peroxy radical 7 [3.25, 3.26].

3) β fragmentation (oxygen abstraction) competes with the capacity of a H donor to form hydroperoxide 3 [3.27].

4) The loss of oxygen from 7 leads to a new carbon W radical 8 with trans,cis stereochemistry [3.21].

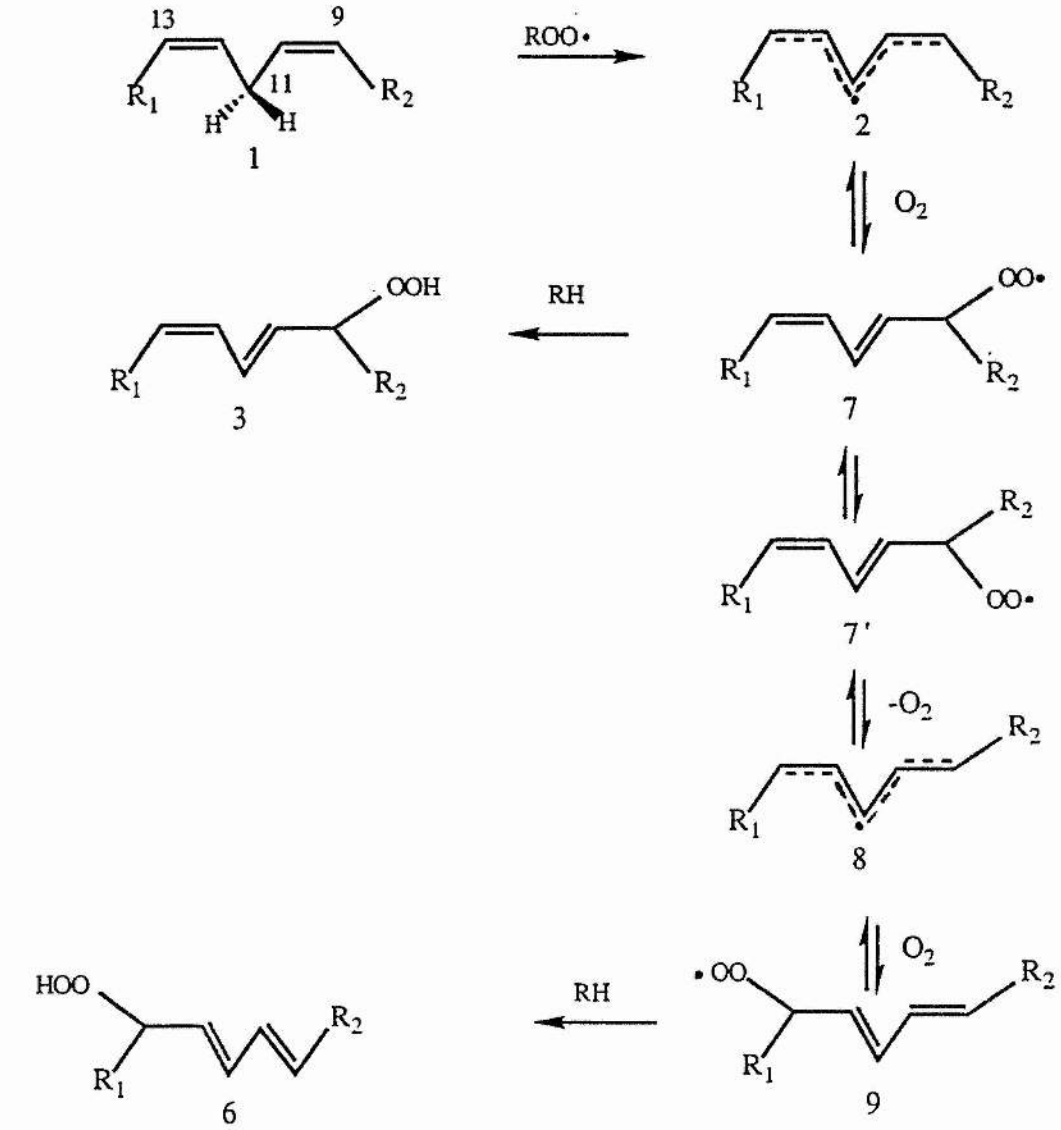
5) A new process of oxygen addition and H-abstraction leads to 6 trans,trans hydroperoxides.

The data concerning linoleate autoxidation are consistent with this mechanism. Thus: 1) the interaction of the unpaired electron on carbon 11 and the π-electrons of adjacent double bonds produces an allylic radical in which the electrons are delocalized over five carbon atoms 2) products are formed from oxygen addition to either end of intermediate bis-allilyc radical and consequently the pseudosymmetry of the system leads to equal amounts of C9 and C13 substituted hydroperoxides. It means that the end carbons of bis-allilyc radical 2 are equivalent. 3) the peroxy radical 7 can react by two alternative pathways: a) it can accept a H-atom from a donor (in this case linoleate) and form the trans,cis hydroperoxide 3; or b) it can isomerize followed by β-fragmentation and formation of trans,trans products. 4) It has been show that at higher temperatures β-fragmentation pathways become competitive with H-atom-transfer reactions and

more trans,trans products (5, 6) are formed (in the case of thermodynamic control); but higher concentration of linoleate lead to more trans,cis products because increasing concentrations of the H-donor favour conversion of 7 to 3 (in the case of kinetic control).

Figure 3.1.1

PORTER' S SCHEME



$R_1 = C_5H_{11}$

$R_2 = (CH_2)_7COOH$

CHAPTER 3.2: SEMIEMPIRICAL CALCULATIONS ON THE LIPID PEROXIDATION PATHWAY

INTRODUCTION

Although a partial confirmation of Porter's scheme has come from the isolation and characterization in chemical model systems, both "in vitro" and "in vivo" [3.20, 3.21, 3.28] of some of the most stable compounds; the less stable intermediates in the scheme have not yet been characterized. This is mainly due to the transient nature of the molecules involved, especially the radicals. However, if one is to fully understand the mechanistic aspects of the reaction pathway, both structural and energetic data on all the molecules involved in the scheme are needed as well as those of the transition states involved.

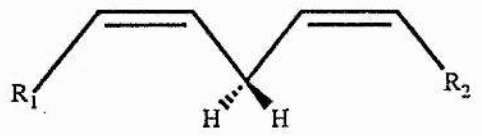
3.2.1) Strategy

In this thesis, Porter's scheme has been studied in some detail by the use of computational chemistry methods [see also ref. 3.29]. In view of the size of the molecules involved, semiempirical methods have been used. The calculations were carried out with either AM1 and the new PM3 Hamiltonians [3.30, 3.31, 3.32], using MOPAC-V5.0 [3.33].

Calculations were performed on Linoleic acid and its derivatives, and on five model compounds containing the key-

functional groups of the original fatty acid involved in the Lipid Peroxidation mechanism.

The model compounds are:



1) $R_1 = CH_3$ $R_2 = CH_3$
2) $R_1 = CH_3$ $R_2 = COOH$
3) $R_1 = CH_3$ $R_2 = CH_2COOH$
4) $R_1 = CH_2CH_3$ $R_2 = COOH$
5) $R_1 = CH_2CH_3$ $R_2 = CH_2COOH$

The theoretical modelling of those reactions in which open-shell molecules are involved would be best done by the UHF formalism. Unfortunately, by this method, an extremely high spin contamination was found in the ground state of all the radical species. Hence, we abandoned this formalism in favour of the RHF-HE method in which the unpaired electron is replaced by two half-electrons of opposite spin. The heats of formation for the closed-shell and open-shell molecules involved in Porter's scheme were all calculated with RHF and RHF-HE methods respectively.

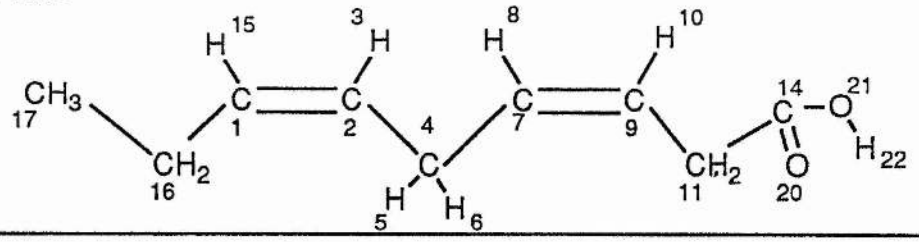
The geometries were optimized tightening all optimization criteria with the PRECISE key-word as implemented in MOPAC, and particularly lowering the norm of the gradient, that is, the magnitude of the vector of derivatives of the energy with respect to the coordinates, to below 0.1 kcal mol^{-1} $Å^{-1}$. The heats of formation are all expressed in Kcal/mol.

3.2.2) Results

In figure 3.2.1, the heats of formation as calculated by AM1, for the compounds involved in the peroxidation of Linoleic acid are reported. The values in brackets refer to PM3 method.

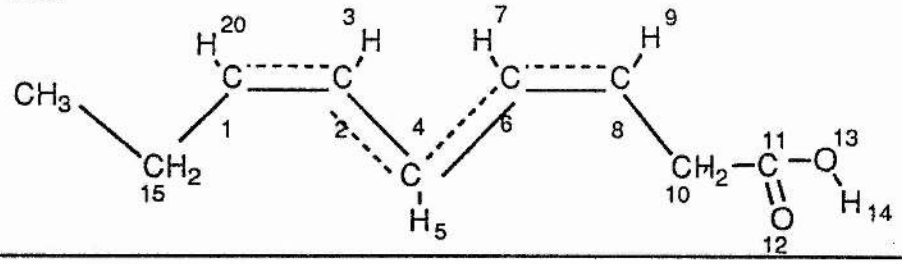
A selected range of internal coordinates of the acid model compound 5 and derivatives are given in table 3.2.1-4. From the reported data it is seen that both methods give similar values.

Table 3.2.1



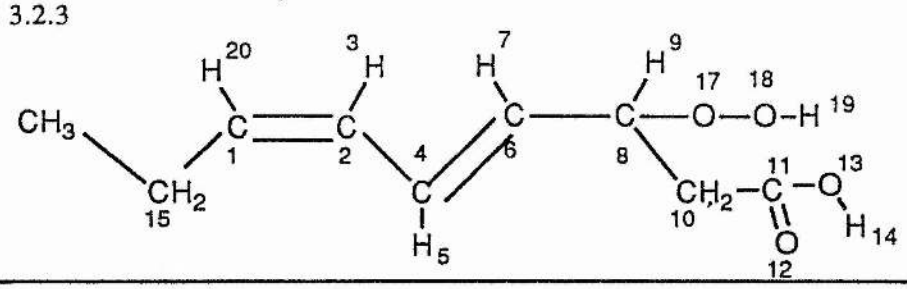
	AM1	PM3
BOND LENGTH (Angstrom)		
C1-C2	1.338	1.336
C2-C4	1.483	1.487
C4-C7	1.483	1.487
C7-C9	1.337	1.336
C9-C11	1.482	1.492
C11-C14	1.497	1.507
C16-C17	1.508	1.513
C16-C1	1.482	1.486
C14-O20	1.234	1.217
C14-O21	1.364	1.357
INTERBOND ANGLES (degrees)		
C1-C2-C3	120.2	119.7
C2-C4-C6	108.9	109.4
C7-C9-C11	124.3	123.1
C16-C1-C2	123.9	124.0
C1-C2-C4	123.4	123.6
C9-C7-C4	124.5	123.9
C1-C16-C17	112.6	112.3
C9-C11-C14	112.4	113.5
O20-C14-O21	116.6	115.2
TORSION ANGLES (degrees)		
C1-C2-C4-C7	176.3	178.7
C2-C4-C7-C9	173.1	179.8
C4-C7-C9-C11	-1.5	-0.03
C7-C9-C11-C14	146.6	179.8
C2-C1-C16-C17	177.5	179.3
C16-C1-C2-C4	-0.3	-0.1

Table 3.2.2



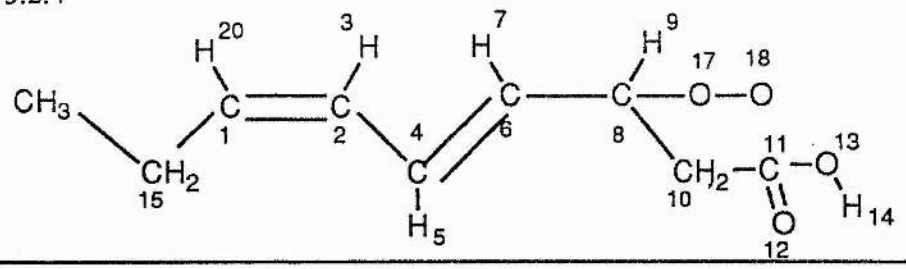
	AM1	PM3
BOND LENGTH (Angstrom)		
C1-C2	1.356	1.355
C2-C4	1.411	1.411
C4-C6	1.413	1.417
C6-C8	1.354	1.352
C8-C10	1.479	1.486
C10-C11	1.597	1.510
C11-O12	1.235	1.128
C11-O13	1.364	1.355
C1-C15	1.479	1.484
C15-C16	1.509	1.514
INTERBOND ANGLES (degrees)		
C1-C2-C4	124.5	124.2
C2-C4-C6	122.2	121.6
C6-C8-C10	124.4	123.8
C8-C10-C11	113.1	110.8
O12-C11-O13	116.4	115.7
C2-C1-C15	124.6	124.7
C1-C15-C16	112.5	112.2
TORSION ANGLES (degrees)		
C1-C2-C4-C6	180.0	179.6
C2-C4-C6-C8	179.9	178.6
C4-C6-C8-C10	-0.7	-1.5
C6-C8-C10-C11	160.3	126.2
C15-C1-C2-C4	0.0	-0.2
C16-C15-C1-C2	-179.2	179.9

Table 3.2.3



	AM1	PM3
BOND LENGTH (Angstrom)		
C1-C2	1.341	1.339
C2-C4	1.448	1.453
C4-C6	1.340	1.338
C6-C8	1.491	1.506
C8-C10	1.526	1.544
C10-C11	1.492	1.506
C11-O12	1.234	1.217
C8-O17	1.461	1.419
O17-O18	1.285	1.512
O18-H19	0.987	0.943
INTERBOND ANGLES (degrees)		
C1-C2-C4	124.6	124.4
C2-C4-C6	122.9	121.7
C6-C8-C10	111.2	108.5
C6-C8-O17	107.2	106.2
C8-C10-C11	113.8	112.4
C8-O17-O18	112.2	107.0
O17-O18-H19	108.1	65.4
O12-C11-O13	116.2	115.8
TORSION ANGLES (degrees)		
C1-C2-C4-C6	179.9	-179.8
C2-C4-C6-C8	179.3	-179.1
C4-C6-C8-C10	69.1	101.9
C6-C8-C10-C11	-161.9	-167.1
C10-C8-C6-O17	117.4	118.7
C6-C8-O17-O18	124.7	141.6
C8-O17-O18-H19	91.0	-37.1

Table 3.2.4



	AM1	PM3
BOND LENGTH (Angstrom)		
C1-C2	1.341	1.340
C2-C4	1.448	1.453
C4-C6	1.340	1.338
C6-C8	1.484	1.496
C8-C10	1.518	1.532
C10-C11	1.495	1.509
C11-O12	1.233	1.217
C8-O17	1.523	1.487
O17-O18	1.160	1.251
INTERBOND ANGLES (degrees)		
C1-C2-C4	124.5	124.4
C2-C4-C6	122.9	121.7
C6-C8-C10	111.6	109.6
C6-C8-O17	106.8	107.0
C8-C10-C11	113.6	112.9
C8-O17-O18	115.8	114.2
O17-O18-H19	78.5	78.0
O12-C11-O13	116.7	116.0
TORSION ANGLES (degrees)		
C1-C2-C4-C6	-177.9	-176.5
C2-C4-C6-C8	-179.3	-179.2
C4-C6-C8-C10	107.1	108.4
C6-C8-C10-C11	-161.0	-168.3
C10-C6-C8-O17	114.4	115.8
C6-C8-O17-O18	108.7	115.1

In table 3.2.5 the heats of formation of the parent acids are reported. The differences in Kcal/mol between the heats of formation calculated by the two methods (Δ A-P) are also reported; the compound 6 is the Linoleic acid.

As far as the parent acid compounds are examined AM1 gives lower heats of formation, and the difference between the two methods increases with the number of alkyl groups of the molecule. This is probably due to a systematic error of AM1 in which the heat of formation of the CH_2 fragment is too negative [3.34].

Table 3.2.5

Acid compounds			
	AM1	PM3	Δ A-P
1	6.58	8.23	-1.65
2	-77.49	-73.60	-3.89
3	-84.20	-79.70	-4.50
4	-84.14	-78.91	-5.23
5	-90.87	-85.07	-5.80
6	-152.75	-133.93	-18.82

If we now consider the heats of formation of the Hydroperoxide derivatives (table 3.2.6):

Table 3.2.6

HYDROPEROXIDES							
	CIS,TRANS				TRANS,TRANS		
	AM1	PM3	Δ A-P		AM1	PM3	Δ A-P
1	-11.21	-15.63	4.42		-12.94	-16.21	4.27
2	-95.23	-95.32	0.1		-95.78	-98.12	2.34

continue

3	-102.48	-102.68	0.2		-103.11	-103.63	0.52
4	-101.92	-100.55	-1.37		-102.03	-102.79	0.76
5	-109.16	-106.55	-2.61		-109.32	-108.30	-1.02
6	-170.45	-156.88	-13.57		-170.65	-157.08	-13.57

we can see that the above discussion is not valid for Hydroperoxides with shorter chains, where PM3 gives lower heats of formation, probably due to the differences between the two methods in dealing with the peroxy bond (CO-OH) [3.34]. In fact from table 3.2.4, and tables 3.2.7 and 3.2.8 it can be seen that the peroxy bond length calculated by AM1 is too short as compared to the available experimental data. PM3 corrected this error but as far as the heats of formation are concerned there are still differences between calculated and experimental values (table 3.2.9).

In table 3.2.7 the peroxy bond calculated by PM3 and AM1 on Hydrogen Peroxide is compared with the experimental value.

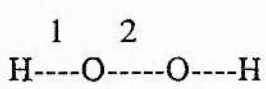


Table 3.2.7 BOND LENGTH (Å)

	1	2
EXP	0.950	1.475
PM3	0.945	1.482
AM1	0.983	1.300

The average peroxy bond calculated on the molecules under examination is reported in table 3.2.8

1 2 3
RC----O----O----H

Table 3.2.8 BOND LENGTH (Å)

	1	2	3
PM3	1.420	1.515	0.943
AM1	1.462	1.287	0.985

In table 3.2.9 the difference between calculated and experimental heats of formation for some molecules in which the peroxy bond is involved are shown.

Table 3.2.9 HEATS OF FORMATION (Kcal/mol)

	Experimental	Δ PM3-Exp	Δ AM1-Exp
H_2O_2	-32.5	-8.3	-2.8
Dimethylperoxide	-30.1	-4.0	3.1
Diethylperoxide	-46.1	6.1	7.7

Here we can see that PM3 gives always lower heats of formation than AM1.

In the following tables 3.2.10 and 3.2.11 the heats of formation for the five model compounds and derivatives are reported. In the first row, the numbers refer to the molecules in figure 3.1.1. In the first column the numbers refer to the model compounds.

Heats of Formation in Kcal/mol for 5 model compounds

Table 3.2.10 AM1

	1	2	7	7'	3	8	9	6
1	6.58	30.75	10.52	10.97	-11.21	29.72	9.37	-12.94
2	-77.49	-53.40	-68.18	-67.46	-95.23	-53.72	-73.16	-95.78
3	-84.20	-59.85	-78.35	-77.80	-102.48	-60.70	-80.75	-103.11
4	-84.14	-60.17	-74.88	-74.16	-101.92	-60.42	-79.47	-102.03
5	-90.87	-66.53	-85.03	-84.60	-109.16	-67.39	-87.04	-109.32

Table 3.2.11 PM3

	1	2	7	7'	3	8	9	6
1	8.24	32.89	22.85	22.44	-15.63	31.53	22.44	-16.21
2	-73.60	-48.73	-53.91	-52.08	-95.32	-50.44	-58.78	-98.12
3	-79.70	-55.51	-63.39	-62.15	-102.68	-55.86	-64.68	-103.63
4	-78.91	-54.00	-59.98	-58.01	-100.55	-55.68	-63.47	-102.79
5	-85.07	-60.74	-67.10	-69.93	-106.55	-61.09	-69.36	-108.30

As already indicated in the previous section, a number of open-shell molecules are investigated. For Semiempirical studies of these molecular systems, the best quantum-mechanical approach would be the UHF formalism. In the following tables (3.2.12 and 3.2.13) the heats of formation as calculated with the RHF and RHF-HE, and UHF formalisms for some of the derivatives of the model compound 5, involved in Porter's scheme, are reported. As we expected for open-shell compounds (2, 7, 8 and 9) the UHF method gives lower heats of formation.

Table 3.2.12 AM1

COMPOUND	RHF, RHF-HE	UHF	$\langle S^2 \rangle$
1	-90.86	-90.86	0.0
2	-60.74	-79.03	1.152
7	-85.03	-86.82	1.045
3	-109.16	-109.56	0.306
8	-67.39	-79.93	1.150
9	-87.04	-88.78	1.049
6	-109.32	-109.63	0.282

Table 3.2.13 PM3

COMPOUND	RHF, RHF-HE	UHF	$\langle S^2 \rangle$
1	-85.07	-85.07	0.0
2	-66.53	-73.07	1.161
7	-67.10	-68.54	0.977
3	-106.55	-106.73	0.228
8	-61.09	-73.42	1.161
9	-69.36	-70.73	0.962
6	-108.30	-108.43	0.212

The expectation value of $\langle S^2 \rangle$ as given by the UHF method is 0, as expected, for the compound 1; while its value for the other closed-shell compounds: 3 and 6 is 0.306 and 0.282 with AM1 and 0.228 and 0.212 with PM3 respectively. For the open-shells radicals 2 and 8 and Peroxides 7 and 9 is about 1 with both AM1 and PM3 Hamiltonians. For these open-shell molecules with

one unpaired electron, the exact expectation value of $<S^2>$ should be 0.75. These higher values of $<S^2>$ are due to a contamination of the unrestricted wave function by higher multiplicity components both for open- and closed-shells. These results forced us to abandon the UHF formalism in favour of the RHF, RHF-HE formalisms.

In order to study the energetic details of the whole pathway, the value of the rotational barriers (activation energies) connecting the rotamers of the same Peroxide (7, 7'), are needed. In table 3.2.14 the rotational barriers are reported. These were obtained by taking the torsional angle as reaction coordinate, and then letting the geometry relax to a minimum energy for each value of the coordinate.

Table 3.2.14

PEROXIDES						
	7 cis-trans		Ea		7' cis,trans	
	AM1	PM3	AM1	PM3	AM1	PM3
1	10.52	22.85	0.73	1.22	10.97	22.44
2	-68.18	-53.91	1.21	0.41	-67.46	-52.08
3	-78.35	-63.39	0.75	0.16	-77.88	-62.15
4	-74.88	-59.98	1.25	0.26	-74.16	-58.01
5	-85.03	-67.10	0.63	0.19	-84.60	-65.93

An interesting observation regarding the Peroxides is the relatively lower heats of formation found in the trans,trans compounds compared with those found in the cis,trans analogues (table 3.2.15). This is particularly true when the 2 and 4 models are examined.

Table 3.2.15

PEROXIDES						
	AM1			PM3		
	trans,trans	cis,trans	Δ t,t-c,t	trans,trans	cis,trans	Δ t,t-c,t
1	9.37	10.52	-1.15	22.44	22.85	-0.41
2	-73.16	-68.18	-4.98	-58.78	-53.91	-4.87
3	-80.75	-78.35	-2.40	-64.70	-63.39	-1.31
4	-79.47	-74.88	-4.59	-63.47	-59.98	-3.49
5	-87.04	-85.03	-2.01	-69.36	-67.10	-2.26
6	-149.40	-148.76	-0.64	-118.13	-117.27	-0.86

In the analysis of the eigenvectors of these open-shell molecules it has been found that for 2 and 4 models, one of the molecular orbitals is consistently more delocalized in the trans,trans peroxides rather than in the cis.trans ones. This is mainly due to the participation, in the trans,trans compounds of 2 and 4 models, of the adjacent double bond C=O of the acid group in extending this orbital. A consequence of this could be that the odd electron can delocalize its charge over the whole molecule and, thereby, better stabilize the trans,trans Peroxides than the cis,trans ones.

3.2.3) Conclusions

In the light of these partial results in which the transition states are not examined, we can consider now the reaction pathway as reported in figure 3.2.1 and its energetic details for all the model compounds here examined. The following conclusions can be made:

The heats of formation of the trans,trans products 6 and cis,trans products 3 for all the model compounds, reported in table 3.2.16, clearly indicate that the trans,trans products are slightly more stable.

Table 3.2.16

HYDROPEROXIDES						
	AM1			PM3		
	trans,trans	cis,trans	Δ t,t-c,t	trans,trans	cis,trans	Δ t,t-c,t
1	-12.94	-11.21	-1.73	-16.12	-15.63	-0.49
2	-95.78	-95.23	-0.55	-98.12	-95.32	-2.80
3	-103.11	-102.48	-0.63	-103.63	-102.68	-0.95
4	-102.03	-101.92	-0.11	-102.79	-100.55	-2.24
5	-109.32	-109.16	-0.16	-108.30	-106.55	-1.75
6	-170.65	-170.45	-0.20	-157.08	-156.88	-0.20

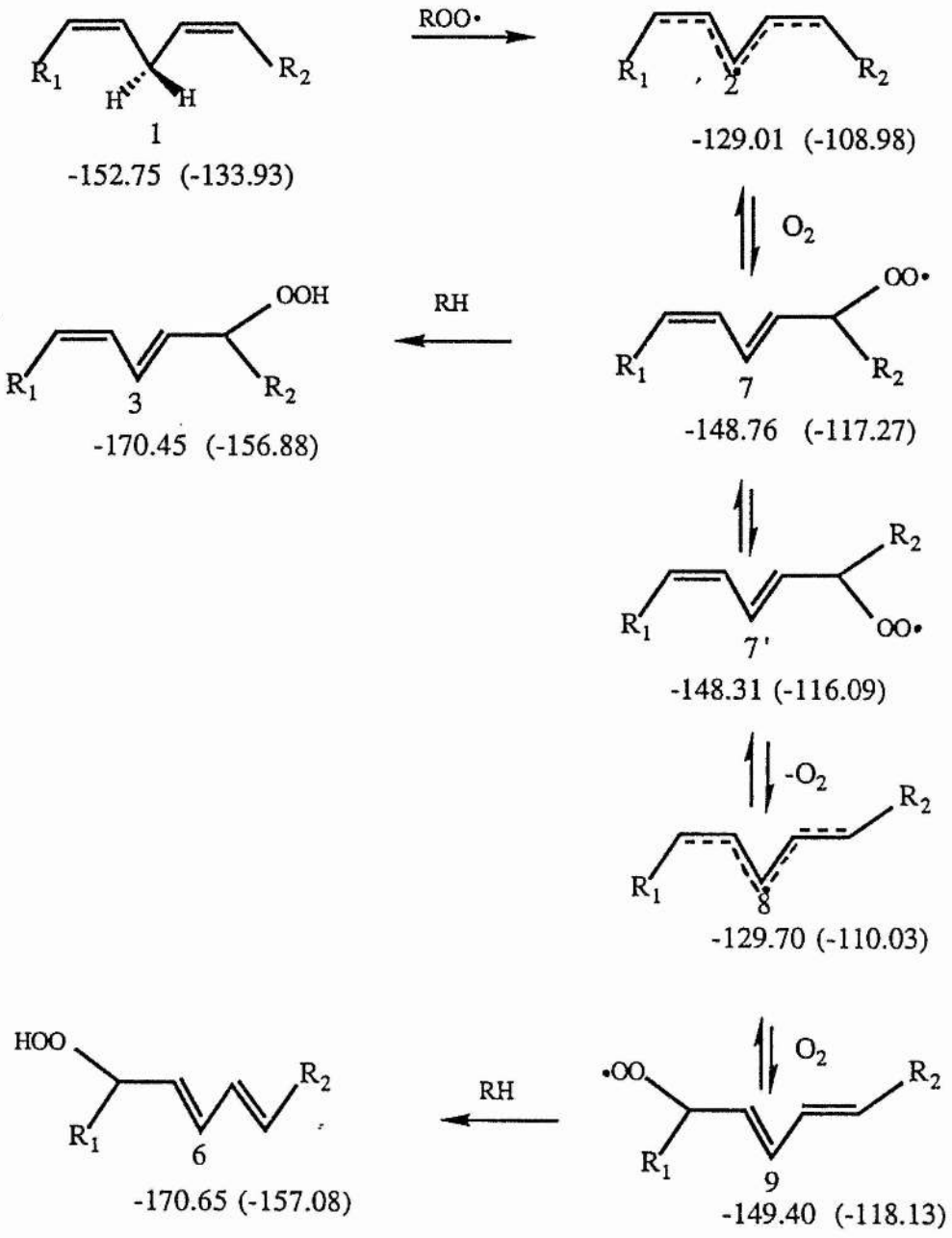
However the energy differences between the two products (6 and 3) are by no means sufficient to overcome the energy gap separating radicals 7' and 8 in the β-fragmentation pathway. This means that the trans,cis products which are formed via a fewer number of transition states, are favoured in comparison with the

trans,trans compounds [kinetic control]. On the other hand tables 3.2.10, 3.2.11 also show that if sufficient energy were given to the system to overcome the barrier to the transition state connecting radicals 7' and 8, the β-fragmentation pathway should be promoted, thus leading to the final trans,trans hydroperoxide 6. Thus an increase in temperature would lead to the most stable dienyl radical 8, and from this to the trans,trans Hydroperoxide compound [thermodynamic control]. These findings agree with the experimental observation that in linoleate autoxidation at high temperatures, more trans,trans product is formed [3.21].

A more correct analysis of the whole pathway should take into account the energetic details of the transition states connecting the various steps of the reaction. Unfortunately, after several attempts, I was unable to characterize, to isolate and to calculate the heats of formation of the transition states involved in this reaction pathway. Nevertheless the above results can be considered exhaustive if we consider that: 1) the transition states connecting compounds 1 and 2 and compounds 2 and 7 are a common way for the formation of both Hydroperoxides; 2) it has been shown experimentally that molecular oxygen abstraction (e.g. from Peroxide 7' to radical 8) and addition (e.g. from compound 8 to Peroxide 9) involved in this reaction pathway are extremely fast processes and so very low activation barriers must be expected. Furthermore the activation barriers of the final stages which involves the stabilization of radicals (from Peroxide

7 to Hydroperoxide 3 and from Peroxide 9 to Hydroperoxide 6) which are geometrical isomers are expected to be very much the same.

FIGURE 3.2.1



$R_1 = C_5H_{11}$, $R_2 = (CH_2)_7COOH$

3.3) REFERENCES

3.1) Bolland J.L.; Quart. Rev. Chem. Soc.; **3**,1 (1949)

3.2) Frankel E.N.; Chemistry and Physics of Lipids; **44,** 73 (1987).

3.3) Chan H.W.S., Autoxidation of unsaturated lipids; Academic Press. New York (1987)

3.4) Slater T.S., Nutrition; 46 (1987).

3.5) Benedetti A., Barbieri L., Ferrali M., Casini A.P., Fulceri R. and Comporti M.; Chem. Biol. Interact.; **35** 331 (1981).

3.6) Fujimoto K., Neff W.E. and Frankel E.N.; Biochem. Biophys. Acta; **795**, 100 (1984).

3.7) MacGregor J.T., Wilson R.E. Neff W.E. and Frankel E.N. Food Chem. Toxicol.; **23** 1041 (1985).

3.8) Mukai F.H. and Goldenstein B.D.; Science; **191**, 868 (1976).

3.9) Munkres K.D.; Mech Age Dev.; **10**, 249 (1979).

3.10) Shamberger R.J. Andreono T.L. and Wills C.E.; J. Natl. Cancer Inst.; **53** 1771 (1974).

3.11) Summerfield F.W. and Tappel A.L.; Analyt. Biochem.; **143** 265 (1984)

3.12) Player T.J. in "Free radical, Lipid Peroxidation and Cancer"; (eds. D.C.H. McBrien and T.F. Slater)173-195 Academic. Press London (1982).

3.13) Barber A. and Bernheim F. in "Advances in Gerontological Research " (ed. B.L. Strehler) Vol II 355-403 Acad. Press New York (1967).

3.14) Arneson R.M., Aloyo V.J., Germain G.S. and Chenevey J.E; Lipids; **13** 383-39 (1978).

3.15) Cheeseman K.H., Collins M., Proudfoot K., Slater T.F., Burton G.W., Webb A.C. and Ingold K.H.; Biochem. J.; **235** 507-514 (1986).

3.16) Cheeseman K.H., Emery S., Maddix S.P., Slater T.F., Burton G.W., and Ingold K.H.; Biochem. J.; **250** 247-252 (1988)

3.17) Cheeseman K.H. Collins M. Maddix S.P., Milia A. Proudfoot K. Slater T.F., Burton G.W., Weeb A.C. and Ingold K.H.; FEBS letters; **209** 191-196 (1986).

3.18) Dahle L.K., Hill E.G. and Holman R.T.; Arch. Biochem. Biophys.; **98** 253 (1962).

3.19) Chan H.W.S., and Levett G.; Lipids; **12** 99 (1977).

3.20) Porter N.A., Lehman L.S. Weber B.A. and Smith K.J.; J. Am. Chem. Soc.; **103** 6447 (1981).

3.21) Porter N.A.; Acts. Chem Res.; **19** 262 (1986)

3.22) Nakano M., Noguchi T., Sugioka K., Fukuyama H., Sato M., Shimizu Y., Tsuji Y. And Inaba H.; J. Biol. Chem.; **250** 2404 (1975).

3.23) Baird M.B., Massie H.R., Pickielniak M.J.; Chem Biol. Interactions; **16** 145 (1977).

3.24) Bascetta E., Gunstone F.D., Walton J.C.; J. Chem Soc. Perkin Trans 1; 603 (1983).

3.25) Chan H.W.S., Levett G. and Matthew J.A.; J. Chem. Soc. Chem. Comm. ; 756 (1978)

3.26) Maillard B., Ingold K.U. and Scaiano J.C.; J. Am. Chem. Soc.; **65** 5095 (1983)

3.27) Porter N.A., Weber B.A., Weenen H., Khan J.A.; J. Am. Chem. Soc.; **102** 5597 (1980)

3.28) F.P. Corongiu, G. Poli, K.H. Cheeseman, M.U. Dianzani, T.F. Slater; Chem Biol. Interactions; **59** 147 (1986)

3.29) G. Saba, P. Scano, P.A. Sedda and C. Thomson; J. Mol. Struct. **204** 379 (1990)

3.30) M.J.S. Dewar, E.G. Zoebish, E.F. Healy, J.J.P. Stewart; J. Am. Chem. Soc. **107** 3902 (1985)

3.31) J.J.P. Stewart; J. Comp. Chem. **10** 209 (1989)

3.32) J.J.P. Stewart; J. Comp. Chem. **10** 221 (1989)

3.33) MOPAC-V5.0; J.J.P. Stewart, (private communication);

3.34) J.J.P. Stewart; J. of Computer-Aided Molecular Design **4** 1 (1990)

CHAPTER 4

STUDIES ON N-METHYL FORMAMIDE

INTRODUCTION

N-Methylformamide (NMF) is a polar organic compound which has been widely used as a solvent. Its antitumor activity against murine tumors was firstly showed by Clarke et al. in 1953 [4.1], when they evaluated a series of formamides as possible solvents for parental administration of organic chemicals in the murine sarcoma 180 screening program. The compounds were found to be too toxic to be used for this purpose; however two agents, namely NMF and Formamide (F), were observed to inhibit the growth of the tumor. In 1956 at the Memorial Sloan-Kettering Cancer Center, Myers et al. conducted the first clinical trial of NMF in the treatment of patients with advanced cancer but symptoms of severe hepatotoxicity caused further studies to be abandoned [4.2].

Interest in NMF was renewed in recent years when activity against human tumor xenografts in nude mice was demonstrated in an NCI (National Cancer Institute) screening program [4.3]. Clinical trials were started in 1982. Since their initiation, extensive work has been devoted to increasing the understanding of the pharmacology, biochemistry, and range of activity of NMF.

CHAPTER 4.1: BIOLOGICAL BACKGROUND

4.1.1) Mechanism of action

In spite of its simple structure and the extensive work carried out on it, the mechanism of action of NMF is still unknown.

NMF has been shown to induce differentiation, maturation (and consequentially less malignancy) in cultured human breast and human colon carcinoma cells [4.4, 4.5, 4.6, 4.7].

NMF has also been shown *in vivo* to have antineoplastic activity [4.8, 4.9, 4.10, 4.11, 4.12]. It is also known that NMF induces radiosensitization and chemosensitization [4.13].

Many preclinical studies with NMF and other polar compounds focused on defining the mechanism of action of these compounds as differentiating agents [4.14]. The induction of cellular differentiation appears to involve multiple metabolic cell events that culminate in terminal differentiation [4.4]. It has been found that NMF induces stabilisation of the membranes and decreases membrane fluidity [4.5]. Other investigators have reported that the metabolic events induced by polar solvents include: (1) an alteration in the transport of Ca^{2+} cations required for cellular commitment to differentiation [4.6]; (2) regulation of transcription of genes that regulate cellular differentiation [4.7] and; (3) modulation of the expression of certain proto-oncogenes (

such as c-*myc*), whose corresponding proteins may be involved in the regulation of differentiation [4.15, 4.16].

Studies by Cordeiro and Savarese indicated that the induction of differentiation and cytostatic effects of NMF may be related to cellular depletion of Glutathione (GSH) levels [4.17]. Glutathione has been implicated in protecting cells against the toxic effects of the antineoplastic agents and/or radiation by both detoxifying the cell and effecting cellular repair induced by these treatments.

Structural requirements for activity are stringent. Studies on N-Methylacetamide (NMA), Formamide (F), N-Ethyl-Formamide (NEF) , N,N-diMethyl-Formamide (DMF), and other Alkylformamides confirmed the requirement for an N-Methyl group and a Formyl group in the molecule for antitumor activity *in vivo*. Interestingly, most of the Alkylformamides were found to be able to induce differentiation in human and murine leukemia cells [4.18, 4.19].

4.1.2) PRECLINICAL ANTITUMOR ACTIVITY

NMF was reported to be active against a number of human tumor xenografts and murine tumors [4.8], activity not showed by other polar solvents. The results of an NCI screening programme showed that there was better than 90% tumor growth inhibition

against human lung, colon, and mammary tumor xenografts implanted beneath the renal capsule [4.3, 4.9, 4.10].

NMF induces terminal differentiation *in vitro* in cultured murine and leukemia cells [4.20, 4.21, 4.22]. After exposure to a differentiating agent, these cells become functionally mature. Cell lines derived from solid tumors after treatment with NMF develop a more benign phenotype characterised by reduced tumorigenicity, loss of clonogenicity in soft agar, and alteration in cellular morphology.

NMF enhanced the radiation sensitivity of tumor cells in vitro and in vivo [4.23].

4.1.3) PRECLINICAL TOXICOLOGY

Preclinical toxicology studies were conducted in mice, rats, and beagle dogs [4.11, 4.12]. The major organ-specific toxic effects noted in all species involved the hepatic system [4.24]. Hepatotoxicity was found to be dose related and reversible in mice. NMF is not directly hepatotoxic, but requires metabolic activation by the host. NMF has been demonstrate to cause depletion of the hepatic glutathione pool (GSH-GSSG), and to be able to bind covalently to hepatic proteins. This is specific to NMF when compared with other formamide derivatives, such as N-hydroxymethyl-Formamide (NMF-OH), NEF and F [4.18, 4.25].
NMF is not a myelosuppressive agent [4.3].

4.1.4) PHARMACOKINETICS

Following IP administration of the radiolabeled drug to mice, radioactivity was found to be freely distributed in all organs. Renal excretion was the primary route of elimination. Besides carbon dioxide (eliminated in the breath) and the unchanged drug, three major urinary metabolites have been detected: methylamine (MA), NMF-OH, and S(N-methyl-carbamoyl)-N-acethylcysteine [4.26, 4.27, 4.28, 4.29]; and one biliary metabolite: S(N-methyl-carbamoyl)glutathione [4.28, 4.30].

4.1.5) CLINICAL TRIALS

Despite the broad range of antineoplastic activities, that have been demonstrated for NMF in diverse preclinical tumor models, and the lack of myelosuppresive activity, its utility as a chemotherapeutic agent has fallen short of original expectations. Phase I and Phase II clinical trials have been conducted with both i.v. and orally administered drug [4.3, 4.31-34]. Virtually no antitumor activity has been seen to date in these trials. In addition, patients treated with NMF have experienced significant gastrointestinal (nausea, vomiting) toxicity and hepatotoxicity [4.3, 4.31-34]. The pharmacological data suggest that NMF concentration administered was insufficient to induce significant cytotoxic effects, differentiating activity or enhancement of radiosensitivity and chemosensitivity in patients with cancer.

The discrepancy in the preclinical and clinical results with this drug may be due to the inability to achieve drug levels in humans comparable to the *in vitro* levels at which the interesting biologic properties of NMF were first observed. Higher drug levels in humans would be associated with intolerable hepatic, gastrointestinal, and constitutional toxic effects. For these reasons a more detailed knowledge of the metabolism of NMF is required [4.2, 4.3, 4.31-34].

4.1.6) METABOLISM

Although much previous work has been dedicated to exploring the metabolism of NMF, the mechanism of its toxicity has not been still elucidated, partly due to the lack of appreciable biotransformation *in vitro*. In fact, the differentiating activity of NMF *in vitro* is probably as the native compound. *In vivo* , NMF is metabolically activated. What is puzzling is the great selectivity of this molecule; it has absolutely no activity on the bone marrow. Therefore the problem with the activity of NMF *in vivo* is really understanding which molecule or metabolite is responsible for the selective mechanism of toxicity [4.35].

Several findings confirm that hepatotoxicity and GSH depletion are strongly related [4.36, 4.37]. NMF is thought to be metabolised to an intermediate which reacts with GSH. Conjugation of GSH with a reactive electrophilic metabolite of NMF

seems to be likely the cause of the GSH depletion, as a gluthatione conjugate of NMF; S(N-Methylcarbamoyl)gluthatione has been identified as a biliary metabolite in humans, rats, and mice [4.28, 4.30]. GSH depletion is specific to NMF when compared with other formamide derivatives, such as NMF-OH, NEF and F which are not hepatotoxic and do not deplete hepatic GSH [4.9, 4.38].

One of the major metabolites of NMF is NMF-OH. It has been showed that NMF-OH has no antitumor activity, that it is not the reactive metabolite of NMF nor the precursor of reactive species responsible for the antitumor activity or hepatotoxicity of NMF [4.12, 4.38].

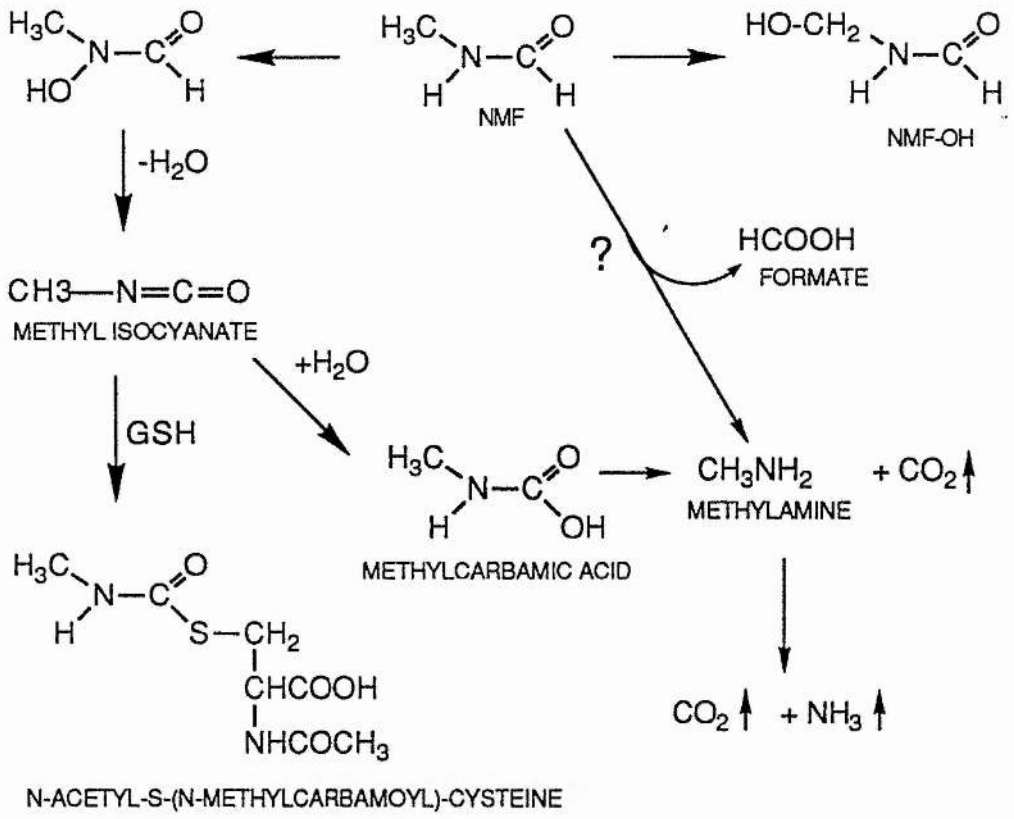
A possible candidate compound for the reactive electrophilic precursor of N-Acetylcysteinyl conjugate of NMF is Methyl Isocyanate [4.26, 4.28]. A number of Isothiocyanates, including Methyl Isocyanate, have been shown to produce analogous N-Acetylcysteinyl conjugates as major urinary metabolites in rats [4.39, 4.40]. Isocyanates are good carbamoylating agents, and will covalently bind to nucleophilic sites in macromolecules such as thiol, hydroxyl and amino groups, heterocyclic nitrogen atoms and carboxylate anion [4.30]. Furthermore, Isocyanates have been shown to interfere with RNA processing, DNA polymerase activity, and other aspects of nucleic acid metabolism [4.42, 4.43].

Methylamine, an endogenous constituent of urine, was detected as a major urinary metabolite of NMF [4.26, 4.27]. Recent

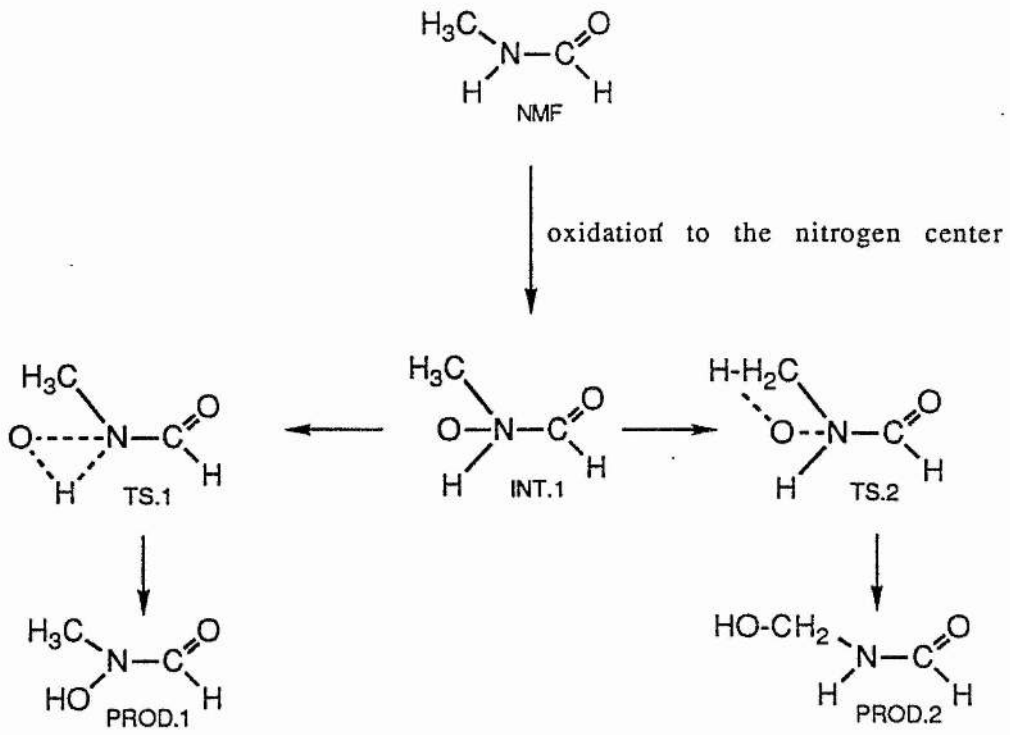
data suggest that one of the possible pathways for Methylamine production is via the breakdown of the postulated reactive metabolite, Methylisocyanate, following reaction with water [4.28].

In summary, it has been demonstrated that: 1) NMF requires metabolic activation by the host; 2) that NMF-OH is not a reactive metabolite of NMF but it is likely a product of a detoxification route; 3) that a metabolite of NMF is able to deplete hepatic glutathione and to covalently bind to nucleophilic sites in macromolecules.

In the light of these and other results, Tulip et al. [4.28] suggested the following metabolic pathway in which Methylisocyanate has been proposed as the reactive metabolite:



In the initial stage NMF can follow two different metabolic pathways: 1) formation of the inactive metabolite NMF-OH. This is likely a detoxification route; 2) formation, through an intermediate, of Methyl Isocyanate, an extremely toxic compound. Both of them start with the oxidation at the nitrogen center of NMF by singlet oxygen. Hence the crucial point of the metabolism of NMF is identified in this initial stage :



Here in the first step, NMF is oxidized by atomic oxygen. Since the metabolization of this molecule takes place mainly in the liver (like most of the exogenous compounds) , the reaction is probably catalysed by hepatic mono-oxygenases (such as Cytochrome P450) [4.29]. In the second step, the N-oxide intermediate after rearrangement yields either N-hydroxy-N-MethylFormamide or N-hydroxyMethyl-Formamide. The nature of the transition states and intermediate are needed in order to study the energetic details of the whole mechanism that cannot be readily characterised by experimental techniques alone. Theoretical studies by quantum-chemical methods can provide this valuable information, as well as estimates of the thermodynamics and kinetics of the reaction.

The aim of this work is to study the structural and energetic data of the compounds involved in this first stage of the metabolism of NMF.

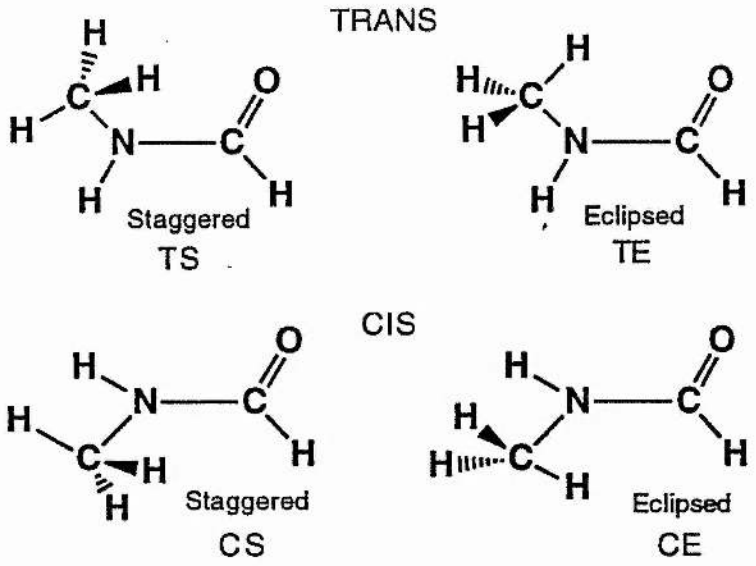
This pathway has been examined using ab initio quantum chemical techniques, employing 3-21G, 6-31G and 6-31G** basis sets. Complete geometry optimisations of all the most stable conformations of NMF, various intermediates, transition states, and products were performed. Energetic details of this reaction have been determined both in gas-phase and in solution by the use of similar methodologies which allow a direct comparison of the gas phase and solution data.

CHAPTER 4.2: AB INITIO STUDIES ON THE STRUCTURE OF NMF IN VACUO AND IN SOLUTION

4.2.1) PROPERTIES OF NMF IN VACUO

Ab initio calculations have been carried out on N-MethylFormamide. The structure of this molecule has been the focus of considerable interest. Studies by X-ray diffraction [4.44], electron diffraction studies [4.45], microwave [4.46], infrared spectroscopy [4.47, 4.48] and neutron diffraction [4.49] have been carried out previously by experimentalists.

For this molecule, there may be trans (methyl group cis to the carbonyl oxygen atom) or cis (methyl group trans to the carbonyl oxygen atom) conformations about the N-C bond and, for either of such conformations, the methyl hydrogens can eclipsed the adjacent N-C bond, or be symmetrically staggered about it. These conformers are illustrated below.



There is a consensus of opinion that the trans isomer is the most stable conformation and that the rotation around N-C bond is severely hindered.

In this work geometry optimizations have been performed from a starting conformer in the vicinity of each of the above four conformations. The calculations have been carried out with basis sets of increasing size, i.e.: STO-3G, 3-21G, 6-31G, 6-31G*, and 6-31G**.

In table 4.2.1 the total energies in Hartrees as calculated with different basis sets for the four conformations are reported.

table 4.2.1

	TS	TE	CS	CE
STO-3G	-205.267684	-205.268527	-205.267200	-205.268034
3-21G	-206.799111	-206.799538	-206.796113	-206.797118
6-31G	-207.865434	-207.865328	-207.862534	-207.863406

continue

6-31G*	-207.961349	-207.961284	-207.958450	-207.959611
6-31G**	-207.971641	-207.971637	-207.968757	-207.969928

The analysis of the total energies shown that with STO-3G and 3-21G, the TE conformation is the most stable, while with the other basis sets, the most stable conformation is the TS. For the trans conformation, the staggered-eclipsed energy difference (Δ S-E in Kcal/mol) has been found to be very small. For the cis conformation, however, the conformer with eclipsed methyl group was calculated to be the more stable with all the basis sets (table 4.2.2).

table 4.2.2

	STO-3G	3-21G	6-31G	6-31G*	6-31G**
TRANS ΔS-E	0.53	0.27	-0.07	-0.04	0.00
CIS ΔS-E	0.52	0.63	0.55	0.54	0.73

The differences expressed in Kcal/mol (and in KJ/mol) between the most stable trans and the most stable cis conformations (Δ T-C) for all the basis sets are reported in table 4.2.3.

table 4.2.3

	STO-3G	3-21G	6-31G	6-31G*	6-31G**
ΔT-C	-0.31(-1.30)	-1.52(-6.36)	-1.27(-5.31)	-1.09(-4.56)	-1.07(-4.48)

In all cases, except for the very small value calculated using the STO-3G basis, the computed energy differences are in good

agreement with the experimental values of -1.6 Kcal/mol [4.50] and -1.44, -1.66 kcal/mol [4.51]. Radom et al. [4.52] calculated the trans-cis difference to be -0.31 Kcal/mol and -1.31 Kcal/mol with STO-3G and 4-31G basis sets respectively;, in either trans or cis conformations the more stable conformation of methyl group was eclipsed. The trans-cis energy difference has been calculated by Sugawara et al. [4.50] to be -1.4 Kcal/mol with a 4-31G basis set. There does not appear to be any experimental information concerning the rotation of the methyl group in NMF.

The results of geometry optimizations of the most stable trans conformations are given in table 4.2.4 . The structure of the trans form determined in the electron diffraction study [4.45] and in the microwave spectroscopy [4.46] are given in the last columns of table 4.2.4.

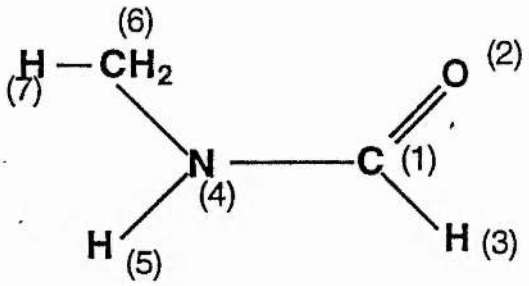


table 4.2 .4

•	STO-3G	3-21G	6-31G	6-31G*	6-31G**	El. diff.[a]	M.W.
R_{12}	1.218	1.215	1.223	1.196	1.196	1.219(5)	1.19
R_{13}	1.105	1.084	1.082	1.091	1.093	1.125[b]	1.09
R_{14}	1.405	1.351	1.346	1.345	1.344	1.366(8)	1.38
R_{45}	1.017	0.996	0.991	0.993	0.992	1.027[b]	1.00
R_{46}	1.461	1.463	1.455	1.448	1.448	1.459(6)	1.47
R_{67}	1.090	1.080	1.080	1.082	1.082	1.114(25)	1.09
α_{213}	124.3	122.2	121.4	122.2	122.3		
α_{214}	124.6	125.2	124.3	124.7	124.7	124.6(5)	
α_{314}	111.2	112.5	114.3	113.0	113.0	112.7[b]	
α_{645}	118.8	118.9	119.2	119.6	119.7		
α_{145}	118.6	119.3	119.0	118.7	118.7	118.7[b]	
α_{146}	122.6	121.7	121.8	121.7	121.6	121.4(9)	
τ_{2145}	180.0	180.0	180.0	179.8	179.8		

R=bond lengths in Å, α=bond angles in degrees, t=torsional angles in degrees. a) standard deviation in parentheses. b) assumed to be equal to the corresponding bond length or bond angle in Formamide

The peptide linkage O=C-N-H is found to be planar in all cases. In the above table 4.2.4, we see that the distances involving the hydrogen atoms are always shorter than those determined

with electron diffraction, but are in good agreement with the microwave data. This may due to the fact that this type of distance is difficult to determine by electron diffraction because of low weighting of the hydrogen. The bonds at the nitrogen atom are in a planar arrangement.

From the overall structural results shown in table 4.2.4, we observe good agreement between the calculated geometries and the experimental data. A comparison of the basis sets shows that the bond length values as computed by STO-3G are always longer; a small split valence basis set such as 3-21G represents a marked improvement over the STO-3G basis set in predicting geometrical structures closer to bigger basis sets with polarization functions such as 6-31G**.

The atomic populations of the trans-staggered conformations as predicted by STO-3G, 3-21G, 6-31G, 6-31G* and 6-31G** basis sets are reported in table 4.2.5. Here we see that results using the STO-3G basis set gives smaller atomic populations in absolute value than the other basis sets.

table 4.2.5

ATOMIC POPULATIONS (in electrons)									
	QC1	QO2	QH3	QN4	QH5	QC6	QH7	QH8	QH9
STO-3G	0.25	-0.27	0.05	-0.37	0.19	-0.07	0.07	0.07	0.07
3-21G	0.63	-0.61	0.19	-0.89	0.36	-0.34	0.20	0.20	0.23

continue

6-31G	0.55	-0.59	0.17	-0.84	0.38	-0.20	0.19	0.19	0.16
6-31G*	0.51	-0.57	0.15	-0.75	0.39	-0.27	0.19	0.19	0.19
6-31G**	0.56	-0.57	0.10	-0.68	0.31	-0.12	0.12	0.14	0.14

In table 4.2.6 structural data for the four conformations of NMF as calculated using the 6-31G** basis set are reported. The trans-cis differences in the structural parameters are not so drastic, in general, but two interesting findings are described here. First, the N-H (R_{45}) bond length in the cis form is slightly longer, and the $\alpha 145$ angle smaller. Secondly, the cis $\alpha 146$ angle is greater by a few degrees than the trans $\alpha 146$ angle. These trans-cis structural differences indicate that there may be an attractive force between the carbonyl group and the methyl group (trans form), and between the carbonyl group and the N-H group (cis form). This can explain the relative stability of the cis form, thereby the small energy difference between trans and cis conformations.

table 4.2.6

6-31G**				
	TS	TE	CS	CE
R_{12}	1.196	1.196	1.195	1.196
R_{13}	1.093	1.092	1.094	1.093
R_{14}	1.344	1.346	1.346	1.347
R_{45}	0.992	0.993	0.995	0.996
R_{46}	1.448	1.447	1.446	1.442
R_{67}	1.081	1.085	1.081	1.085
α_{213}	122.3	122.0	122.1	122.3
α_{214}	124.7	125.7	125.0	124.9
α_{645}	119.7	118.9	119.8	119.3
α_{145}	118.7	117.8	116.0	115.5
α_{146}	121.6	123.3	124.4	125.5
τ_{2145}	179.9	180.0	180.0	180.0
τ_{7645}	0.7	61.0	-0.5	60.3

R=bond lengths in Å, α=bond angles in degrees, t=torsional angles in degrees.

In table 4.2.7 the total dipole moments in Debye calculated by different basis sets are reported. The experimental value in gas phase has been found to be 3.81 Debye [4.53].

table 4.2.7

TOTAL DIPOLE MOMENT (Debye)					
	STO-3G	3-21G	6-31G	6-31G*	6-31G**
TRANS STAG	2.69	4.10	4.57	4.13	4.12
TRANS ECL	2.56	3.98	4.47	4.04	4.03
CIS STAG	2.68	4.40	4.87	4.31	4.32
CIS ECL	2.65	4.38	4.87	4.29	4.30

4.2.2) Properties of NMF in solution

A better description of properties of molecules involved in biological systems should take into account solvent effects. For this purpose the method of Tomasi et al. [4.54, 4.55] has been chosen here to simulate the presence of the solvent (water in this case).

The numerical results presented in this section refer to the NMF molecule, kept at fixed geometries and surrounded by a medium of dielectric permittivity ε=78.8. The calculations have been performed on the four most stable conformations of NMF i.e.: trans-staggered, trans-eclipsed, cis-staggered and cis-eclipsed, with STO-3G, 3-21G, 6-31G, 6-31G* and 6-31G** basis sets. Unfortunately no optimization of geometries was available within this method, hence the geometries used are the most stable ones as calculated in vacuo with the pertinent basis set. The analysis is extended to solvent effects on the MO energies, on the dipole moments and on the atomic populations.

The shape of the cavity is formed by 9 interlocking spheres centered on the atoms, each having the radius of the size of the pertinent van der Waals value. In the case of trans-eclipsed and cis-eclipsed geometries, an extra sphere has been added in order to eliminate the outer spaces in which a molecule of solute of finite size can not penetrate. We remark here that the number of the spheres and the values of the radii are the unique empirical parameters introduced in the computational procedure.

The average total surface and the average total volume of the cavity for the different conformations as computed by different basis sets are 85.26 $Å^2$ and 55.10 $Å^3$ respectively. The surface have been divided in tesserae, the average number is: 373.55.

The free energies G (in Hartrees) of the molecule in solution (see Chapter 1) calculated for the most stable conformations of NMF with different basis sets are reported in table 4.2.8.

table 4.2.8

FREE ENERGIES IN SOLUTION (Hartrees)				
	TS	TE	CS	CE
STO-3G	-205.282239	-205.282292	-205.282155	-205.282463
3-21G	-206.830616	-206.830000	-206.829410	-206.829584
6-31G	-207.901030	-207.900010	-207.900306	-207.900309
6-31G*	-207.989273	-207.988750	-207.988415	-207.988916
6-31G**	-208.000787	-207.999393	-207.999088	-207.999604

The solvation free energies ΔG (i.e. the differences in Kcal/mol (KJ/mol) between the free energies in solution and the total energies in vacuo ($\Delta G=G-E^0$)) are reported in table 4.2.9.

table 4.2.9

	ΔG (TS)	ΔG (TE)	ΔG (CS)	ΔG (CE)
STO-3G	-9.13 (-38.20)	-8.64 (-36.15)	-9.38 (-39.25)	-9.05 (-37.86)
3-21G	-19.77 (-82.72)	-19.12 (-80.00)	-20.89 (-87.40)	-20.37 (-85.23)
6-31G	-22.34 (-93.47)	-19.12 (-91.06)	-23.70 (-99.17)	-23.16 (-96.89)
6-31G*	-17.52 (-73.30)	-17.24 (-72.11)	-18.80 (-78.67)	-18.39 (-76.94)
6-31G**	-17.74 (-74.22)	-17.42 (-72.87)	-19.04 (-79.66)	-18.63 (-77.96)

For comparison with table 4.2.3, the differences of free energies in solution in Kcal/mol (KJ/mol) between trans and cis conformations are shown in table 4.2.10 .

table 4.2.10

	STO-3G	3-21G	6-31G	6-31G*	6-31G**
Δ T-C	0.11 (0.45)	-0.65 (-2.71)	-0.45 (-1.89)	-0.22 (-0.92)	-0.18 (-0.75)

The STO-3G basis set, in contrast with experimental results, gives the cis conformation as the most stable in solution. The free energy difference between trans and cis conformations in solution (trans the most stable) decreases (except that for STO-3G basis set), with the increasing number of basis functions, showing the same trend as in vacuo (see table 4.2.3). The analysis of the trans, cis energies in vacuo and in solution shows that the trans-cis difference is reduced by the solvent.

The dipole moments (in Debye) in solution and their differences with in vacuo values, are reported in tables 4.2.11 and 4.2.12 respectively.

table 4.2.11

TOTAL DIPOLE MOMENT IN SOLUTION (debyes)					
	STO-3G	3-21G	6-31G	6-31G*	6-31G**
TRANS STAG	3.62	5.94	6.68	6.01	6.10
TRANS ECL	3.51	5.89	6.67	6.00	6.04
CIS STAG	3.71	6.46	7.20	6.36	6.41
CIS ECL	3.65	6.40	7.15	6.31	6.37

table 4.2.12

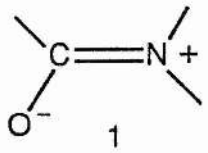
Δ TOTAL DIPOLE MOMENT IN SOLUTION-IN VACUO (DEBYES)					
	STO-3G	3-21G	6-31G	6-31G*	6-31G**
TRANS STAG	0.93	1.84	2.11	1.88	1.98
TRANS ECL	0.95	1.91	2.20	1.96	2.01
CIS STAG	1.03	2.06	2.33	2.05	2.09
CIS ECL	1.00	2.02	2.28	2.02	2.07

The atomic populations calculated in vacuo and in solution as computed using the 6-31G** basis set are shown in table 4.2.13.

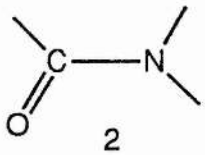
table 4.2.13

6-31G** ATOMIC POPULATION (in electrons)								
	IN VACUO				IN SOLUTION			
	TS	TE	CS	CE	TS	TE	CS	CE
Q C1	0.56	0.57	0.58	0.58	0.62	0.62	0.64	0.64
Q O2	-0.57	-0.58	-0.58	-0.58	-0.77	-0.77	-0.78	-0.78
Q H3	0.10	0.10	0.09	0.09	0.15	0.15	0.15	0.14
Q N4	-0.68	-0.68	-0.68	-0.68	-0.65	-0.66	-0.66	-0.65
Q H5	0.31	0.30	0.32	0.32	0.34	0.34	0.33	0.33
Q C6	-0.12	-0.14	-0.13	-0.13	-0.11	-0.11	-0.11	-0.11
Q H7	0.12	0.19	-0.14	0.13	0.14	0.14	0.15	0.14
Q H8, H9	0.14	0.12	0.13	0.13	0.14	0.14	0.14	0.14

After solvation, a decrease of the electronic population can be seen on all the hydrogen atoms. At the carbonyl group there is a considerable increase in the electronic population on the oxygen atom in solution (with respect to the in vacuo value), accompanied by a decrease of electronic population in the C1 atom. The solvent effect seems to give more weight to a polar structure, C^{+}-O^{-} in solution than in vacuo. These findings are in agreement with the increase of the total dipole moments in solution. These data, considering also the decrease of the atomic population in N4 atom, are in agreement with the intuitive view that for NMF, solvation emphases the contribution of the ionic structure (1) with respect to the covalent one (2).



1



2

In table 4.2.14 the MO energies in Hartrees for HOMO and LUMO orbitals calculated with the 6-31G** basis set are reported. The HOMO-LUMO gaps in NMF decrease in going from the vacuo situation to solution, due to a lowering of the LUMO energies. This fact may indicate a higher reactivity of the molecule in solution.

table 4.2.14

6-31G** HOMO-LUMO (in eV)						
	IN VACUO			IN SOLUTION		
	HOMO	LUMO	Δ H-L	HOMO	LUMO	Δ H-L
TRANS STAG	-0.391	0.209	-0.600	-0.391	0.199	-0.590
TRANS ECL	-0.390	0.213	-0.603	-0.391	0.203	-0.594
CIS STAG	-0.389	0.200	-0.589	-0.389	0.193	-0.582
CIS ECL	-0.389	0.213	-0.602	-0.390	0.201	-0.591

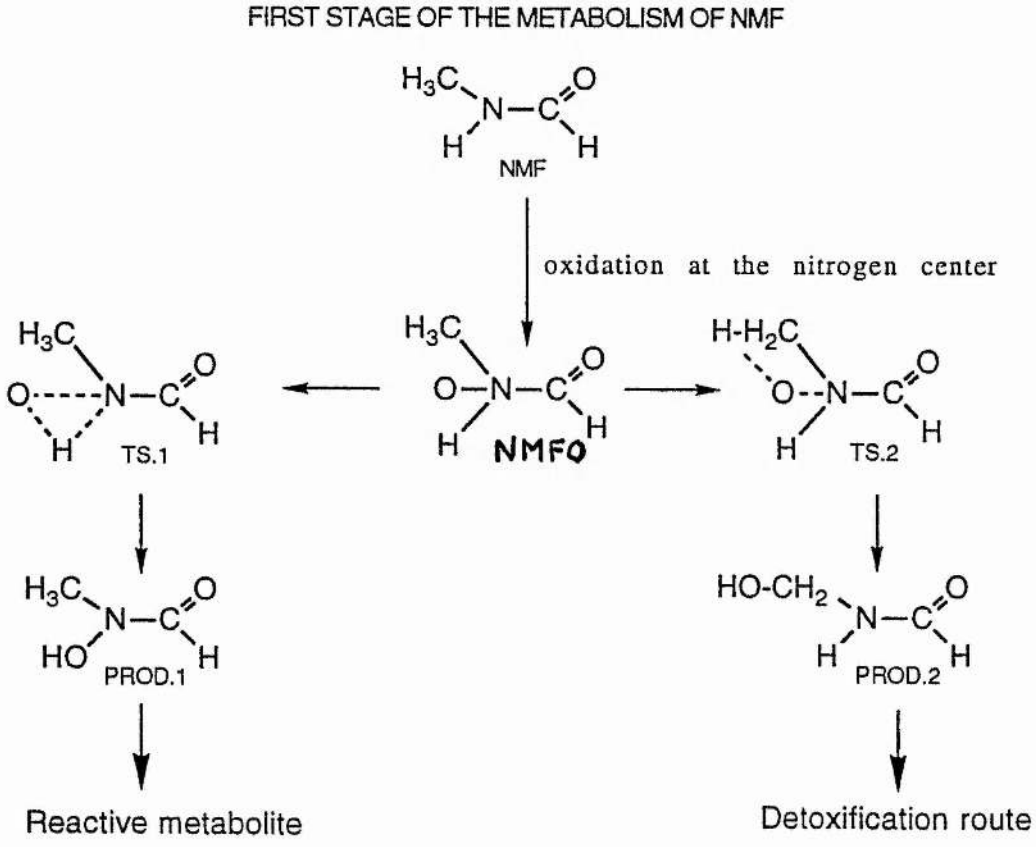
It must be pointed out that the implementation of the continuum solvation model does not provide, at the present stage, for the optimization of the geometry in solution; hence, particular attention must be paid when comparing results.

CHAPTER 4.3: AB INITIO STUDIES ON THE METABOLISM OF NMF

INTRODUCTION

As stated in chapter 4.1, the metabolism, of NMF can follow two different pathways. One is a detoxification route that results in the formation of the inactive metabolite NMF-OH. The other route leads to the reactive metabolite methyl-isocyanate through the formation of an intermediate [4.28].

The crucial point of the metabolism of NMF has been identified in the first stage of the whole pathway in which two different routes compete, thus leading to completely different biological events.



In order to explore these two alternatives, the energetic details of this first stage of the whole pathway have been studied by means of ab initio methods.

Complete geometry optimization of all the stationary points: intermediate (NMFO), transition states (TS.1, TS.2), and products (PROD.1, PROD.2) were performed. The reaction potential energy surface has been studied at different level of accuracy using the 3-21G, 6-31G, and 6-31G** basis sets.

4.3.1) INTERMEDIATE (NMFO)

The first proposed step is an enzymatic oxidation at the nitrogen center of NMF by active oxygen. Since this is a common pathway to the formation of both products, its energetic details are not important to our aim. This reaction leads to an intermediate NMFO which rearranges to two different products through different transition states. In table 4.3.1 the geometries of NMFO are reported.

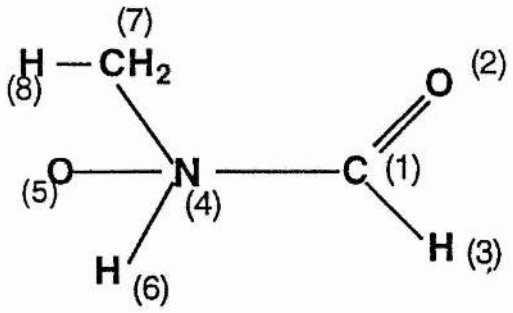


table 4.3.1

	3-21G	6-31G	6-31G**		3-21G	6-31G	631G**
R_{12}	1.195	1.200	1.174	α_{314}	110.9	112.3	110.0
R_{13}	1.073	1.074	1.083	α_{546}	107.2	104.7	109.6
R_{14}	1.438	1.434	1.455	α_{145}	94.0	100.9	102.7
R_{45}	1.552	1.531	1.378	α_{146}	114.6	111.9	109.0
R_{46}	1.006	1.002	1.011	α_{547}	107.1	109.0	111.8
R_{47}	1.479	1.473	1.471	α_{647}	114.1	112.2	109.0
R_{78}	1.080	1.080	1.080	τ_{2147}	-1.7	-0.8	2.0
α_{213}	126.3	124.8	126.7	τ_{2146}	-139.2	-132.0	-120.4
α_{214}	122.7	122.8	123.2	τ_{2145}	109.8	117.1	123.4
α_{147}	117.0	116.8	114.5				

R=bond lengths in Å α=interbond angles in degrees τ=torsional angles in degrees

The most evident difference between the geometries of NMFO as calculated by the three basis sets is the bond length R_{45} between the nitrogen and oxygen atoms. We can see that the 6-31G** basis set result for this distance is 0.174 Å shorter than for the 3-21G basis set, and 0.153 Å shorter than for the 6-31G basis set. The shortening of this bond length due to the inclusion of the polarization functions yields a decrease of bond angles α_{147} and α_{146} and an increase of α_{547} and α_{546}.

The computed total energies are shown in table 4.3.2.

table 4.3.2

TOTAL ENERGIES (HARTREES)			
	3-21G	6-31G	6-31G**
NMFO	-281.127773	-282.574798	-282.709925

4.3.2) PRODUCT 1

Product 1, N-hydroxy N-methyl Formamide, is thought to be the precursor of the reactive highly toxic metabolite N-Methyl Isocyanate. Thus its formation may be responsible, in the final step, for the toxic side effects of treatments with NMF.

In the following tables 4.3.3 and 4.3.4 the geometries and the total energies of PROD1 are reported.

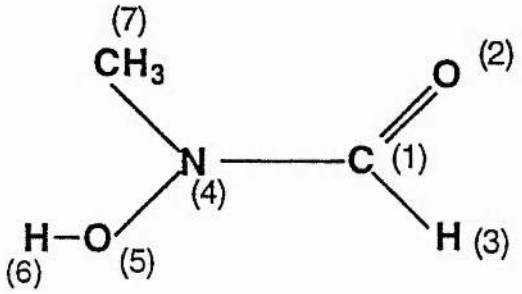


table 4.3.3

	3-21G	6-31G	6-31G**		3-21G	6-31G	631G**
R_{12}	1.209	1.218	1.190	α_{314}	111.6	113.1	111.9
R_{13}	1.079	1.079	1.089	α_{741}	122.3	124.1	120.8
R_{14}	1.371	1.361	1.368	α_{745}	113.0	114.2	113.0
R_{45}	1.437	1.406	1.380	α_{654}	105.1	108.2	106.0
R_{56}	0.969	0.954	0.945	α_{541}	112.1	113.9	112.9
R_{47}	1.462	1.451	1.447	τ_{2145}	162.2	166.0	162.1
R_{78}	1.080	1.080	1.082	τ_{2147}	23.2	18.8	23.8
α_{213}	124.5	123.1	123.9	τ_{6541}	115.5	111.5	115.2
α_{214}	123.9	123.8	124.2				

R=bond lengths in Å α=interbond angles in degrees τ=torsional angles in degrees

Here the N_4-O_5 bond is calculated to be slightly shorter as computed using the 6-31G** basis with respect to the other basis sets here examined.

table 4.3.4

TOTAL ENERGIES (HARTREES)			
	3-21G	6-31G	6-31G**
PROD1	-281.178145	-282.621815	-282.767920

4.3.3) PRODUCT 2

The formation of PROD2, N-Hydroxymethylformamide (NMF-OH), is likely a detoxification route. PROD2 has been found to be an inactive metabolite of NMF. In the following tables 4.3.5 and 4.3.6 the geometries and the total energies of NMF-OH as computed by different basis sets are reported.

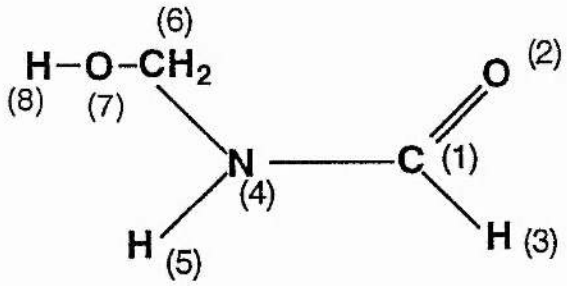


table 4.3.5

	3-21G	6-31G	6-31G**		3-21G	6-31G	631G**
R_{12}	1.220	1.225	1.198	α_{314}	114.4	115.2	113.8
R_{13}	1.082	1.080	1.091	α_{645}	120.2	119.3	119.0
R_{14}	1.349	1.349	1.349	α_{145}	120.5	119.3	117.9
R_{45}	0.997	0.992	0.994	α_{146}	118.5	120.5	119.6
R_{46}	1.463	1.448	1.446	α_{876}	108.5	112.2	108.3
R_{67}	1.420	1.416	1.385	α_{764}	113.1	113.0	113.5
R_{78}	0.969	0.953	0.946	τ_{2146}	0.6	1.1	6.0
α_{213}	122.6	121.6	122.2	τ_{7645}	-111.3	-101.2	-92.7
α_{214}	123.0	123.2	124.0	τ_{8764}	-64.9	-71.6	-70.3

R=bond lengths in Å α=interbond angles in degrees τ=torsional angles in degrees

table 4.3.6

TOTAL ENERGIES (HARTREES)			
	3-21G	6-31G	6-31G**
PROD2	-281.241556	-282.690097	-282.834597

4.3.4) TRANSITION STATE 1

The transition state TS1 connects the intermediate NMFO to the PROD1. Transition structures are normally characterized by weak "partial" bonds, that is those being broken or formed. In this transition state, the N-H bond is being broken and the O-H is being formed. In table 4.3.7 the geometries of TS1 are reported.

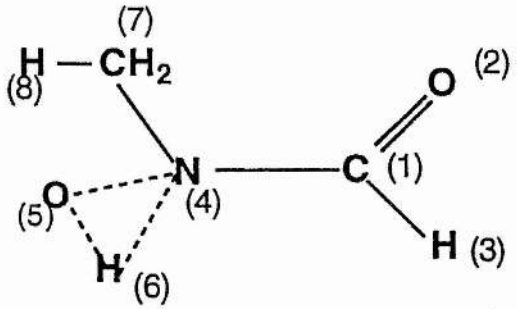


table 4.3.7

	3-21G	6-31G	6-31G**		3-21G	6-31G	631G**
R_{12}	1.203	1.210	1.181	α_{147}	119.7	120.9	119.3
R_{13}	1.077	1.077	1.085	α_{654}	45.2	46.3	45.1
R_{14}	1.393	1.386	1.402	α_{145}	108.5	109.9	109.6
R_{45}	1.626	1.618	1.518	α_{146}	119.8	119.0	118.3
R_{46}	1.157	1.173	1.092	α_{547}	111.0	111.8	112.2
R_{56}	1.216	1.191	1.256	α_{647}	120.5	119.9	121.5
R_{47}	1.465	1.456	1.452	α_{645}	48.3	47.3	54.6
R_{78}	1.080	1.078	1.080	τ_{2147}	16.4	16.0	14.4
α_{213}	124.7	123.5	125.1	τ_{2146}	-162.9	-159.8	-154.8
α_{214}	123.3	123.0	123.0	τ_{2145}	145.2	148.8	145.7

R=bond lengths in Å α=interbond angles in degrees τ=torsional angles in degrees

The total energies are reported in the following table 4.3.8.

table 4.3.8

TOTAL ENERGIES (HARTREES)			
	3-21G	6-31G	6-31G**
TS1	-281.087545	-282.526477	-282.662933

4.3.5) TRANSITION STATE 2

The transition state TS2 connects NMFO with PROD2. Here the H2C-H bond and N-O bonds are broken with a rearrangement in the H2C-O-H bonds. In table 4.3.9 the geometries of TS2 are reported.

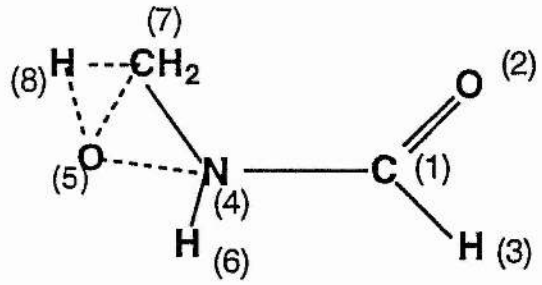


table 4.3.9

	3-21G	6-31G	6-31G**		3-21G	6-31G	631G**
R_{12}	1.207	1.213	1.186	α_{641}	119.4	120.0	116.9
R_{13}	1.077	1.077	1.087	α_{741}	119.0	117.0	119.8
R_{14}	1.389	1.385	1.390	α_{541}	98.4	100.3	106.4
R_{45}	1.956	1.949	1.861	α_{745}	66.8	67.3	66.3
R_{46}	0.999	0.998	0.997	α_{758}	48.9	48.7	50.8
R_{47}	1.468	1.453	1.449	α_{478}	104.5	104.4	104.8
R_{75}	1.938	1.920	1.843	α_{746}	119.0	120.0	117.6
R_{78}	1.520	1.515	1.516	α_{645}	119.4	119.0	118.7
R_{58}	1.714	1.700	1.666	τ_{2145}	67.6	60.3	71.7
α_{213}	124.6	123.8	124.7	τ_{2147}	-0.4	-3.0	0.0
α_{214}	122.3	112.8	112.4	τ_{2146}	-161.7	-160.7	-152.8
α_{314}	112.8	112.6	112.4				

R=bond lengths in Å α=interbond angles in degrees τ=torsional angles in degrees

table 4.3.10

TOTAL ENERGIES (HARTREES)			
	3-21G	6-31G	6-31G**
TS2	-281.007138	-282.445590	-282.542007

4.3.6) RESULTS

In table 4.3.11 the total energies, in Hartrees, of the compounds involved in the first stage of the metabolism of NMF are reported.

TABLE 4.3.11

COMPOUND	3-21G	6-31G	6-31G**
NMFO	-281.127773	-282.574798	-282.709925
PROD.1	-281.1767920	-282.621815	-282.767920
PROD.2	-281.241556	-282.690097	-282.834597
TS1	-281.087545	-282.526477	-282.662933
TS2	-281.007138	-282.445590	-282.542007

The activation barriers in Kcal/mol (KJ/mol in brackets) of pathway 1 are reported in table 4.3.12.

table 4.3.12

Activation Barriers	
3-21G	25.242 (105.637)
6-31G	30.320 (126.890)
6-31G**	29.486 (123.400)

Here we can see that the 6-31G basis set gives the higher barrier of activation.

Pathway 2 is an exothermic reaction. The activation barriers in Kcal/mol (KJ/mol) are reported in table 4.3.13.

table 4.3.13

Activation Barriers	
3-21G	75.695 (316.786)
6-31G	81.074 (339.295)
6-31G**	105.364 (440.948)

The activation barriers for the pathway 2 are significantly higher than those of pathway 1. On the other hand the products are much more stable.

Figure 4.3.1 reports the energies of reaction in Hartrees of both pathways, as calculated by 3-21G, 6-31G and 6-31G** basis sets respectively.

In figure 4.3.2 the energies (in KJ/mol) of the reaction pathway as calculated by 6-31G** are reported.

From the overall results the following observations can be done: 1) the activation barriers of pathway 2 are significatively higher than those of pathway 1; 2) PROD2, as calculated by 6-31G** basis set, is 4.094 Kcal/mol more stable than PROD1. This means that in gas phase the formation of PROD1 is favoured when the reaction is under kinetic control. On the other hand if sufficent energy is given to the system to overcome TS2, or better an enzymatic reaction is involved, the formation of the more stable PROD2 is the favoured.

figure4.3.1.

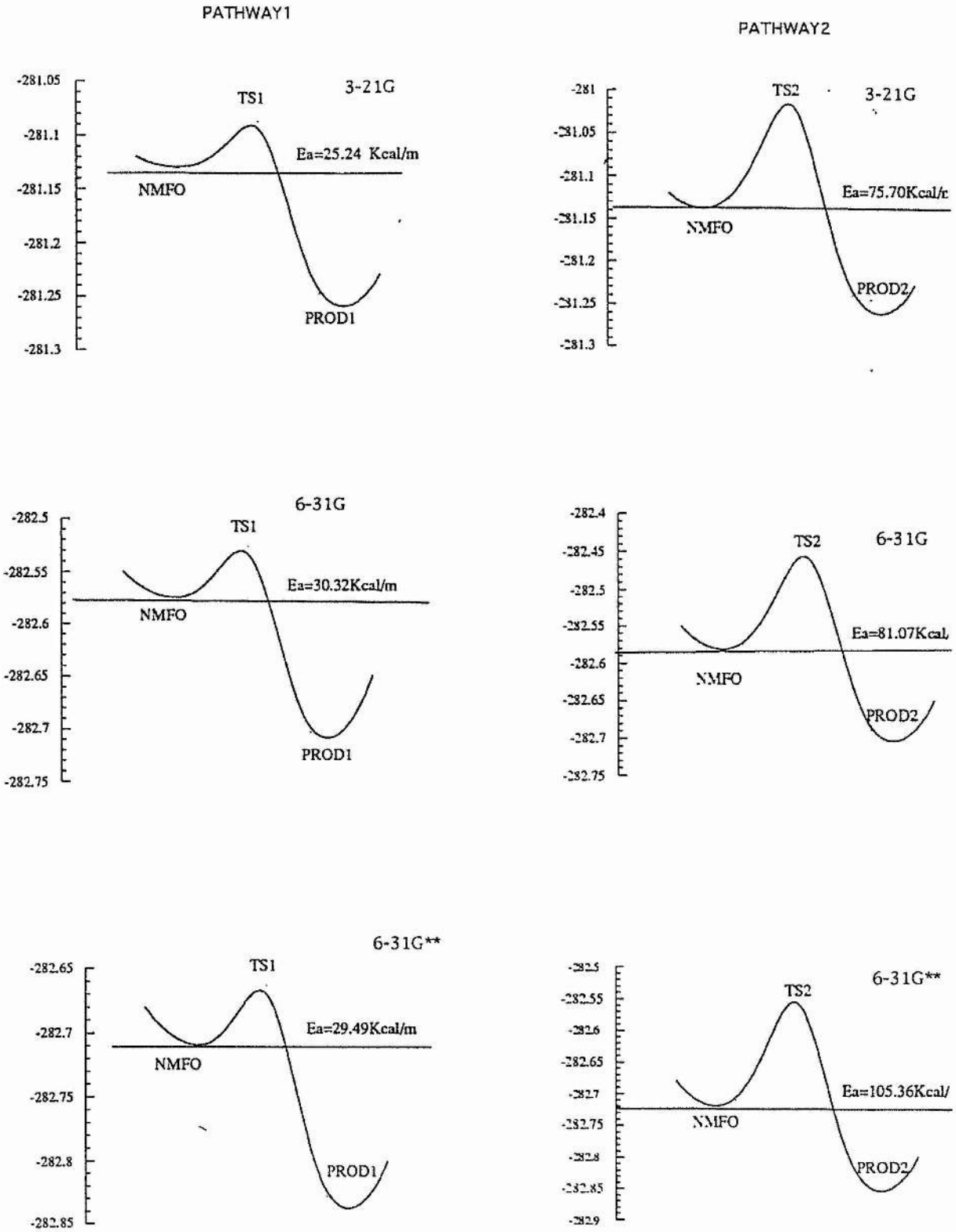
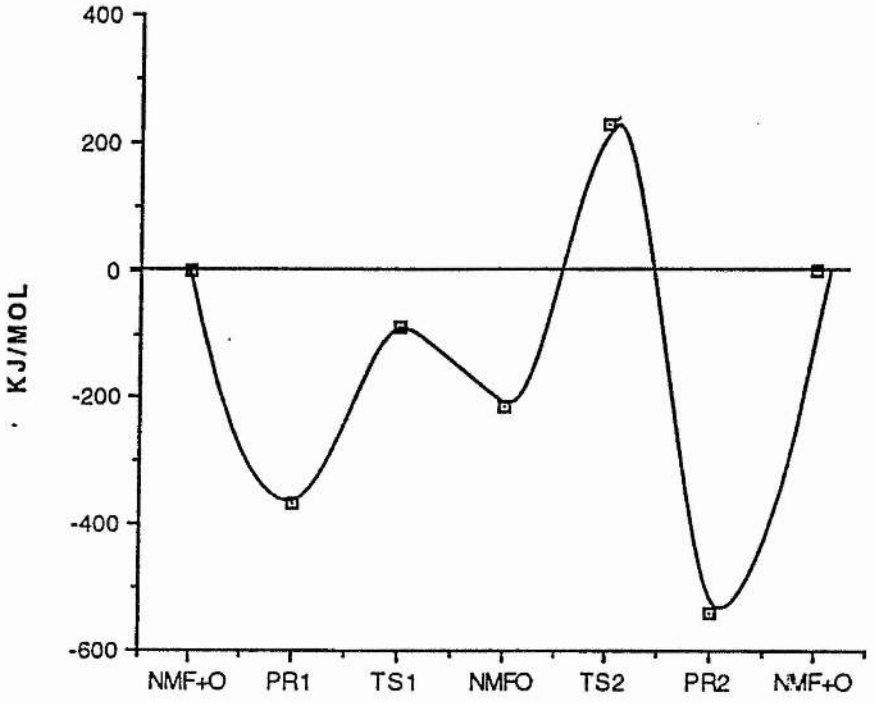

PATHWAY1
PATHWAY2
3-21G
TS1
Ea=25.24 Kcal/m
NMFO
PROD1
-281.05
-281.1
-281.15
-281.2
-281.25
-281.3
TS2
Ea=75.70Kcal/r
PROD2
-281
-281.05
-281.1
-281.15
-281.2
-281.25
-281.3
6-31G
TS1
Ea=30.32Kcal/m
NMFO
PROD1
-282.5
-282.55
-282.6
-282.65
-282.7
-282.75
TS2
Ea=81.07Kcal/
PROD2
-282.4
-282.45
-282.5
-282.55
-282.6
-282.65
-282.7
-282.75
6-31G**
TS1
Ea=29.49Kcal/m
NMFO
PROD1
-282.65
-282.7
-282.75
-282.8
-282.85
TS2
Ea=105.36Kcal/
PROD2
-282.5
-282.55
-282.6
-282.65
-282.7
-282.75
-282.8
-282.85
-282.9

figure 4.3.2



The sum of the total energy of NMF and oxygen atom has been taken as the point at 0 energy.

CHAPTER 4.4: AB INITIO STUDIES ON THE METABOLISM OF NMF IN SOLUTION

INTRODUCTION

The energetic details of the potential energy surface of the first stage of the metabolism of NMF have been examined in presence of a solvent. The method used here to simulate the presence of a solvent is the continuum approach by Tomasi et al. [4.54, 4.55]. The solvent is represented by a continuum having a dielectric constant of 78.5. The effect of the solvent effects on equilibrium structures, transition states through the connecting pathways has been studied. Solvation of the fixed geometries of the gas phase ground states has been investigated. The free energies (G), quantities directly comparable with the total energies in gas phase (E^0), have been determined.

4.4.1) Results

In the following tables 4.4.1-5 we report free energy (G) in Hartrees and difference between free energy and total energy in the gas phase (E^0) in Kcal/mol for all the molecules involved in the first stage of the metabolism of NMF.

table 4.4.1

NMFO		
	G	G-E^0
3-21G	-281.172006	-27.76
6-31	-282.602294	-17.25
6-31G**	-282.756268	-29.08

table 4.4.2

PROD1		
	G	G-E^0
3-21G	-281.209798	-19.86
6-31G	-282.658034	-22.73
6-31G**	-282.794454	-16.65

table 4.4.3

PROD2		
	G	G-E^0
3-21G	-282.277551	-22.59
6-31G	-282.730253	-25.20
6-31G**	-282.863865	-18.37

table 4.4.4

TS1		
	G	G-E^0
3-21G	-281.118420	-19.37
6-31G	-282.560129	-21.12
6-31G**	-282.692286	-18.42

table 4.4.5

TS2		
	G	$G-E^0$
3-21G	-281.039982	-20.61
6-31G	-282.481153	-22.32
6-31G**	-282.572285	-19.00

The last columns of the above tables show that all the compounds here examined are more stable in solution than in vacuo phase; and that the energy differences between in solution and in vacuo states are (apart from NMFO molecule) smaller with 6-31G** basis set and larger with 6-31G basis set.

Table 4.4.6 reports the activation barriers in Kcal/mol (KJ/mol) of pathway 1 in solution as calculated by 3-21G, 6-31G, and 6-31G** basis sets.

table 4.4.6

Activation Barriers	
3-21G	33.63 (140.74)
6-31G	26.45 (110.69)
6-31G**	40.15 (167.97)

Table 4.4.7 reports the activation barriers in Kcal/mol (KJ/mol) of pathway 2 in solution as calculated by 3-21G, 6-31G, and 6-31G** basis sets.

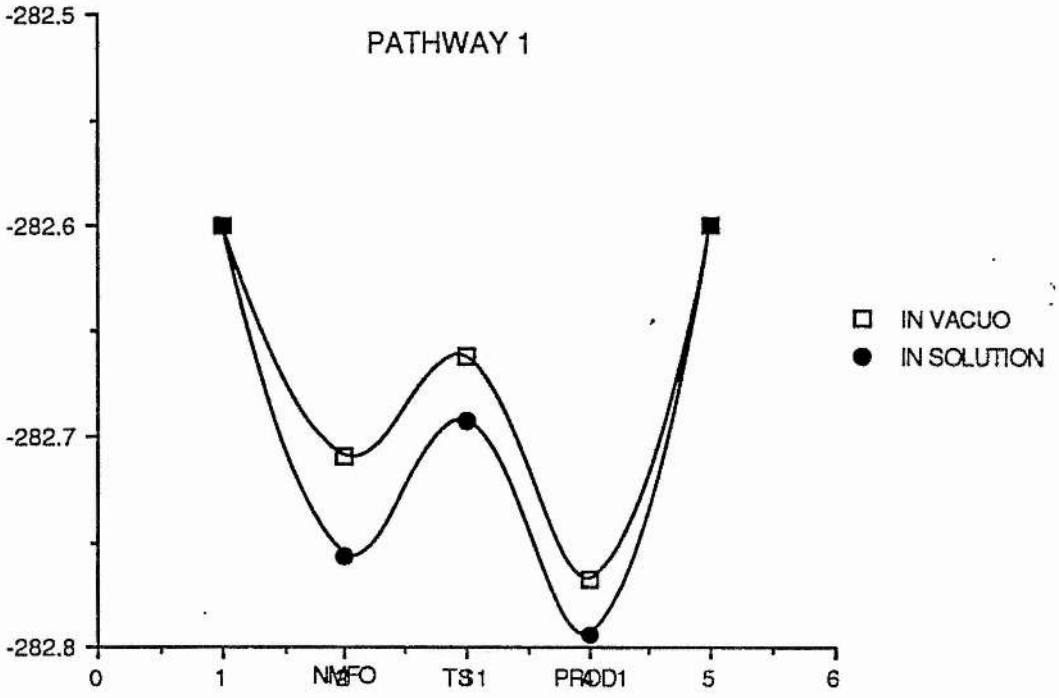
table 4.4.7

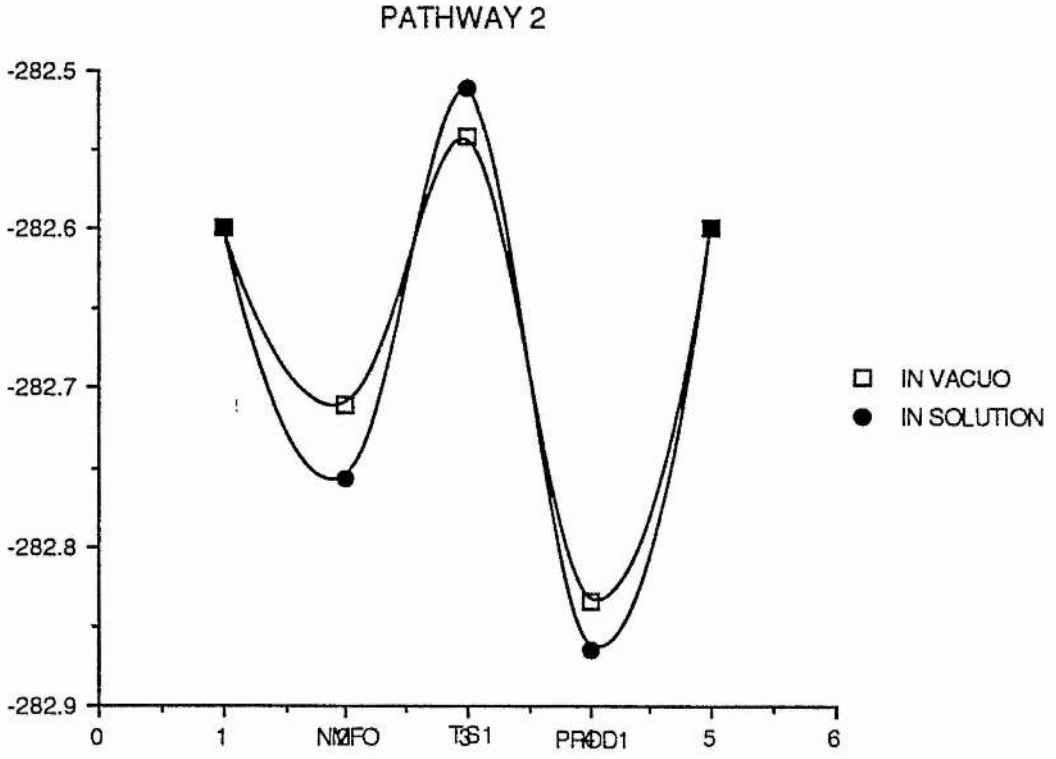
Activation Barriers	
3-21G	50.18(209.95)
6-31G	76.02(318.06)
6-31G**	115.45(483.05)

In the above tables we can see that pathway 2 presents activation barriers higher than pathway 1, showing the same trend as in vacuo.

If we compare these energies with the barriers shown in table 4.3.11 and 4.3.12, we can see that only the 6-31G calculations of pathway 1 give a barrier which is lower than in vacuo. With 3-21G and 6-31G** basis sets, the reactions involved in both pathways seem to be less favorable in solution than are in vacuo.

In figure 4.4.1 pathway 1 and 2 are represented. The data refer to 6-31G** calculations in vacuo and in presence of the solvent.


PATHWAY 1
-282.5
-282.6
-282.7
-282.8
0
1
5
6
IN VACUO
IN SOLUTION


PATHWAY 2
-282.5
-282.6
-282.7
-282.8
-282.9
0
1
5
6
IN VACUO
IN SOLUTION

4.5) References

4.1) D.A. Clarke, F.S. Philips, S.S. Stenberg, R.K. Barclay, C.C. Stock; Proc. Soc. Exp. Biol. Med. **84** 203 (1953)

4.2) W.P.L. Myers, D.A. Karnofsky, J.H. Burchenal; Cancer, **9** 949 (1956)

4.3) National Cancer Institute. Clinical brochure: N-Methylformamide (NMF) NSC 3051. Division of Cancer Treatment, National Cancer Institute, Bethesda, MD. (1982)

4.4) P.A. Marks, R.A. Rifkind; Cancer **54** 2766 (1984)

4.5) G.H. Lyman, H.D. Preisler; Nature **262** 360 (1976)

4.6) L.F. Chapman; Dev. Biol. **79** 243 (1980)

4.7) W.H. Straling; Nucleic Acids Res **3** 1203 (1976)

4.8) M. Iwakawa, P.J. Tofilon, N. Hunter; Clin.Expl. Metastasis **5** 289 (1987)

4.9) A. Gescher, N.W. Gibson, J.A. Hickman; BR. J Cancer **45** 843 (1982)

4.10) S.P. Langdom, A. Gescher, J.A. Hickman; Eur. J. Cancer Clin. Oncol. **20** 699 (1984)

4.11) S.P. Langdon, D. Chubb, A. Gescher; Toxicology **34** 173 (1985)

4.12) H. Whitby, A. Gescher, L. Levy; Biochem. Pharmacol **2** 295 (1984)

4.13) G.Zupi, M. Marangolo, G. Arancia, C. Greco; Cancer Res. **48** 6193 (1988)

4.14) S.P. Langdon and J.A. Hickman; Toxicology **43** 239 (1987)

4.15)D. Chatterjee, A. Mendelsohn, P.R. Shank and T.M. Savarese; Cancer Res. **49** 3910 (1989)

4.16) H.E. Skipper, F.M. Schabel, V. Binns; Cancer Res. **15** 143 (1955)

4.17) R.F. Cordeiro and T.M. Savarese; Cancer Res. **46** 1297 (1986)

4.18) A.J. Shaw, A. Gescher, J. Mraz; Toxicol. Appl. Pharmacol. **95** 162 (1988)

4.19) P. Kestell, M.D. Threadgill and A. Gescher; Drug Metabol Mol. Man 217 (1987)

4.20) M. Tanaka, J. Levy, M. Terada; Proc. Natl. Acad. Sci. USA **72** 1003 (1975)

4.21) H.D. Preisler, G. Lyman; Cell Differ. **4** 179 (1975)

4.22) S.P. Langdon and J.A. Hickman; Cancer Res. **47** 140 (1987)

4.23) D.L. Dexter, E.S. Lu, S.F. Bliven; Cancer Res **44** 4942 (1984)

4.24) H. Whitby, S.B. Chahwala, A. Gescher; Biochem Biophys Res Commun **125** 712 (1984)

4.25) P. Kestell, M.D. Threadgill, A. Gescher, A.P. Gledhill, A.J. Shaw and P.B. Farmer; J. Pharmacol Exp. Ther. **240** 265 (1987)

4.26) P. Kestell A.P. Glehill, M.D. Threadgill and A.Gescher; Biochem. Pharmacol. **35** 2283 (1986)

4.27) P. Kestell, A. Gescher, J.A. Slack; Drug Metab. Dispos. **13** 587 (1985)

4.28) K. Tulip, J.A. Timbrell; Arch. Toxicol. **62** 167 (1988)

4.29) K. Tulip , J.A. Timbrell, J.K. Nicholson, I. Wilson; Drug Metab. Dispos. **14** 746 (1986)

30) M.D.Threadgill, D.B. Axworthy, T.A. Baillie et al.; J. Pharmacol. Exp. Ther. **242** 312 (1987)

4.31) E.K. Rowinsky, D.A. Noe, D.W. Orr, L.B. Grochow, D.S. Ettinger, P.A. Donehower; J. Natl. Cancer Inst. **80** 671 (1988)

4.32) E.N. Spremulli, D.L. Dexter, F.J. Cummings et al.; Proc. Am. Soc. Clin. Oncol. **2** 24 (1983)

4.33) D.S. Ettinger, D.W. Orr, A.P. Rice et al.; Cancer Treat. Rep. **69** 489 (1985)

4.34) J.S.A. Jeffrey, N. Tait, H. Silva; M. Eisenberger; Am. J. Clin. Oncol. **12** 41 (1989)

4.35) J. Hickman: Private Comunication (1989)

4.36) P. G. Pearson, A. Gescher, E.S. Harpur; Biochem. Pharmacol. **36** 381 (1987)

4.37) P.G. Pearson, A. Gescher, E.S. Harpur; Biochem. Pharmacol. **36** 385 (1987)

4.38) P.G. Cooksey, E.N. Gate, A. Gescher, J.A. Hickman, S.P. Langdon; Biochem. Pharmacol. **32** 3037 (1983)

4.39) G. Brusewitz, B.D. Cameron, L.F. Chassaud; Biochem. J. **162** 99 (1977)

4.40) W. H. Mennicke, K. Gorler, H. Krumbiegel; Xenobiotica **13** 203 (1982)

4.41) S. Cohen and E. Oppenheimer; "The chemistry of cyanates" John Wiley New York pp923 (1977)

4.42) B.B. Baril, E.F. Baril, J. Laszlo And G.P. Wheeler; Cancer Res. **35** 1 (1975)

4.43) H.E. Kahn, K.W. Kohn, L. Widerlite and D. Gullion; Cancer Res. **34** 1982 (1974)

4.44) H. Ohtaky, S. Itoh and B.M. Rode; Bull. Chem. Soc. Jap., 59 271 (1986)

4.45) M. Kitano and K. Kuchitsu.; Bull. Chem. Soc. Jap., **47** 631 (1974)

4.46) R.A. Elzaro; Dissertation, University of Ohio (1973)

4.47) S. Ataka, H. Takeuchi and M. Tasumi; J. Molec. Struct., 113 147 (1984)

4.48) I. Suzuki; Bull. Chem. Soc. Jap., **35** 540 (1962)

4.49) J. Neuefeind, P. Chieux and M.D. Zeidler; Molec. Phys. 76 143 (1992)

4.50) Y. Sugawara et al.; Chem. Phys., **62** 339 (1981)

4.51) F.A.L. Anet and M. Squillacote; J. Am. Chem. Soc., **97** 3243 (1975)

4.52) L. Radom and N.V. Riggs; Aust. J. Chem., **35** 1071 (1982)

4.53) R.M. Meighan and R.H. Cole; J. Phys. Chem. **68** 503 (1964)

4.54) S. Miertus, L. Scrocco, and J. Tomasi ; Chem. Phys. **53** 117 (1981).

4.55) S. Miertus and J. Tomasi ; Chem. Phys. **65** 239 (1982).

CHAPTER 5

HARDWARE AND SOFTWARE FACILITIES

Introduction

In this chapter the computer facilities and the programs used in this thesis are reported.

The work reported in this thesis has been done in part at the University of St. Andrews (U.K.) and in part at Universita' degli Studi di Cagliari (Italy), therefore the resources of both Universities are here reported.

Section 5.1 deals with computer hardware. Section 5.2 includes a brief discussion on the computer software.

5.1) Computer Hardware

The University of St. Andrews provided all the students and members of the staff with two VAX II/785 computers, running the VMS operating system. Unfortunately the resources available in terms of disk space and computer memory in the two VAX' s were too small to perform enough calculations on it. The computer facilities also include several Classic Macintoshes and IBM compatible Personal Computers.

Our research group in St. Andrews was provided with Micro VAX II/GPX colour workstation. It runs the VMS operating system and has 13MBytes of RAM It is linked to an FPS-500 (Floating Point System) Mini-supercomputer by Ethernet network. It runs the

UNIX operating system and it is a two processor system: one scalar and one vector. This means that a number of different operations can be performed simultaneously rather than sequentially (as it happens in normal scalar processors), and this greatly increases the computational speed of the programs which run on the vector processor. The FPS-500 has 128-MBytes of memory. The speed of this machine, which belongs to our research group, permitted me to run the calculations concerning Chapter 2 in very short times.

In Cagliari I had the access to a MicroVAX II/GPX VAXStation and to a MicroVAX 3100-40 having 13 Mbytes and 16 Mbytes of memory, respectively. Both machines belong to Dipartimento di Scienze Chimiche. They were linked by Ethernet network with a VAX 6410 of 38MBytes of memory belonging to the Dipartimento di Fisica, where I had access as well. All the machines mentioned above run a VMS operating system.

One Macintosh LC and one Macintosh Classic were available. The Macintoshes installed in both Universities have been utilized to write this thesis up, thank to their powerful word processors and graphic programs.

5.2) Computer Software

The *ab initio* calculations reported in this thesis have been performed with the Gaussian90 and the earlier version Gaussian88 programs [5.1, 5.2]. Gaussian is a connected system of programs for

performing *ab initio* and semiempirical molecular orbital calculations, it is written in Fortran77 language. Because of the amount of memory allocation and of disk space needed to compute and to store the two electron integrals needed, to ameliorate its performance, the program is made to be strictly connected with the characteristic of the machines in which it is installed. Therefore to make any modification is a difficult task.

The program is divided in independent links called one after one by the main route generated by the user. In order to run calculations which include the presence of the solvent (based on the method explained in Chapter 1, section 1.19) using Gaussian programs, we introduced in the latter few routines provided to us by the research group of prof. Tomasi. For the same purpose, a few original links have been modified. In figure 5.1 an example of input for a RHF calculation using STO-3G basis set without geometry optimization is reported.

figure 5.1

```
1     $DEFINE GAUSSR_EXEDIR [USER.SCANO]
2     $RUNGAUSS
3     #RHF/STO-3G
4
5      WATER
6
7     0,1
8     O
9     H 1 0.95
10    H 1 0.95   2 105.5
11
```

The input for the same calculation which include also solvation effects, is reported in figure 5.2.

figure 5.2

```
1       $DEFINE GAUSSR_EXEDIR [USER.SCANO]
2       $RUNGAUSS
3       #NONSTD
4
5         1/29=10000/1;
6         2/10=1,12=2/2;
7         3/5=5,11=1,25=14,30=1/1,2,3,11,14;
8         4/7=1/1;
9         5//1;
10        6/19=1/1,4,5;
11        3/5=5,11=1,25=14,30=1,43=2/1,2,3,11,14;
12        5//1;
13        6//4,5;
14        3/5=5,11=1,25=14,30=1,43=2/1,2,3,11,14;
15        5//1;
16        6//4,5;
17        3/5=5,11=1,25=14,30=1,43=2/1,2,3,11,14;
18        5//1;
19        6//4,5;
20        3/5=5,11=1,25=14,30=1,43=2/1,2,3,11,14;
21        5//1;
22        6/19=1/1;
23        99/5=1,9=1/99;
24
25       WATER
26
27      0,1
28      O
29      H 1 0.95
30      H 1 0.95 2 105.5
```

continue

```
3 1
3 2          3 1 78.5
3 3          1 1.4
3 4          2 1.2
3 5          3 1.2
```

The two inputs are in different format. The first is a standard route; here line 3 includes the keyword commands which specify the quantum mechanical procedure, the basis set and, eventually, the control features involving the various aspects of the execution.

Lines 8, 9 and 10 report the nuclear positions in Z-matrix. Within Z-matrix each atom is related to previously specified atoms by a bond lenght, an angle and a dihedral angle. An alternative way to specify the geometry is by the use of cartesian coordinates.

The format of the figure 5.2 is a Non-Standard Route, generated by the previous standard route and modified in order to include solvation links.

The first line addresses the program to the directory in which the new links and the modified ones are situated. The new links are L604 and L605, they are involved in the computation of the cavity and in the calculation of the MEP in presence of the solvent respectively. The modified links are: L301 in which to the nuclear repulsion energy is added the solvent charge contribution due to the presence of the solvent; L302 where the charges on the surface of the cavity, as previously calculated in L605 are taken into account in the computation of the one-electron integrals; and L501

in which the total and electronic energies in vacuo and in solvation phase are calculated. From line 5 to 23 there are listed the links to be run one after the one. Lines 10, 13, 16 and 19 call the links L604 and L605.

Line 25 is the title and lines 28, 29 and 30 dialogue with the geometry of the molecule in Z-matrix format.

In line 32 the options regarding the solvation are listed: 1) indicates how many spheres to be built around the solute, 2) it is a flag that indicates in which format the spheres are described in the following lines, 3) dielectric of the solvent. Following indication 2 the next lines (33, 34, and 35) describe the characteristic of the cavity surrounding the solute. In this case there are three spheres, the first one is to be built centered on atom one (i.e. oxygen) and has a ray of 1.4 Å (i.e. of the Van der Waals size). The second and the third spheres are centered on the hydrogen atoms and have a ray of 1.2 Å. The other way to specify the characteristic of the spheres is to indicate their cartesian coordinates and their rays; in this case the second number of line 32 should be 2.

The semiempirical calculations reported in this thesis have be run with MOPAC V5.0 [5.3]. MOPAC is a semiempirical molecular orbital program which incorporates many separate functions. The input has the same philosophy of Gaussian program: one line of keyword commands and geometry specification in Z-matrix by default.

Concurrent with the development of MOPAC, a compatible graphics package has been wrote. This program, called DRAW [5.4], acts as a general purpose data manipulation program. It has been very useful for generating and checking the input data for MOPAC. It has been used as the graphical interpretation of the results of MOPAC calculations.

5.3) References

5.1) GAUSSIAN 90; Revision I; M.J. Frish , M. Head-Gordon, G.W. Trucks, J.B. Foresman, H.B. Schlegel, K. Raghavachari, M. Robb, J.S. Binkley, C. Gonzales, D.J. Defrees, D.J. Fox, R.A. Whiteside, R. Seeger, C.F. Melius , J. Baker, R. Martin, R.L. Kahn, J.J.P. Stewart, S. Topiol, and J.A. Pople; Gaussian, Inc.; Pittsburgh, PA; (1990).

5.2) GAUSSIAN 88; M.J. Frish , M. Head-Gordon, H.B. Schlegel, K. Raghavachari, J.S. Binkley, C. Gonzales, D.J. Defrees, D.J. Fox, R.A. Whiteside, R. Seeger, C.F. Melius , J. Baker, R. Martin, R.L. Kahn, J.J.P. Stewart, E.M. Fluder, S. Topiol, and J.A. Pople; Gaussian, Inc., Pittsburgh, PA; (1988).

5.3) J.J.P. Stewart; MOPAC V5.0; private comunication;

5.4) D.M. Storch; DRAW; Quantum Chemistry Program Exchange, No 493, 1985.

CHAPTER 6

CONCLUSIONS

CONCLUSIONS

In this thesis I have applied the Quantum Chemistry methods to novel problems relevant to carcinogenesis mechanisms. The methods chosen are *ab initio* and Semiempirical Molecular Orbitals.

In Chapter 1 a description is given of the basic theory and the methods used, including a discussion of a novel treatment of solvation which I have implemented and tested.

I have examined the validity of the theories used in the study of aromatic heterocyclic compounds in Chapter 2. There experimental values are compared with theoretical results. AM1 and PM3 Hamiltonians have been employed. The results suggest that although the more recent PM3 represents a remarkable improvement on AM1 there are still some discrepancies between calculated and experimental data which can be traced out only when a larger statistics will be available.

The major study of Lipid Peroxidation mechanism is described in Chapter 3. AM1 and PM3 Hamiltonians are used to study in details the scheme proposed by Porter. The theoretical results are in good agreement with experimental findings.

In Chapter 4, the studies of NMF are described. Properties of the molecule, as well as the pathway of reaction proposed, are studied in details with a high degree of accurancy. In that Chapter the solvation continuum model is introduced and tested. In spite of

the drawback of the lack of the optimization of the geometry in solution, the data agree with the results in gas phase and furthermore introduce a closer reproduction of the natural environment of this biological system.

Finally the hardware and software used in this thesis are described in Chapter 5 . There, it is remarked that the computational speed (expecially the parallelization of the computation) and the amount of RAM and of the storage capacity are the major elements to be improved in the molecular modelling of biological systems.